



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 118756

TO: Michael Borin
Location: rem/2a55/2c70
Art Unit: 1631
Tuesday, April 06, 2004

Case Serial Number: 09/997807

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Borin,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

• Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rpm** and **.rpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

STIC-Biotech/ChemLib

118756

From: Borin, Michael
Sent: Monday, April 05, 2004 6:04 PM
To: STIC-Biotech/ChemLib
Subject: Search request:09/997807

Examiner: M.Borin
AU: 1631

Remsen 2A55
Tel.: 20713

RE: 09/997807; peptide polymer

Please conduct search of polypeptide SEQ ID No. 2 against the commercial and interference protein databases.

Thank you

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 4/5/04
Date Completed: 4/6/04
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: osp
WWW/Internet: _____
Other (specify): _____



STIC SEARCH RESULT FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup. Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:24:31 ; Search time 60 Seconds

(without alignments)
974.789 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VXYTLIAIGIITASAAALAL.....EGMLFDSLPIVNFQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1042	100.0	207	5 AAU99735	Pyrodicti
2	1042	100.0	207	7 ADB46044	CanA fusi
3	553.5	53.1	170	5 AAU99736	Pyrodicti
4	553.5	53.1	170	7 ADB46046	CanB fusi
5	540	51.8	178	5 AAU99737	Pyrodicti
6	540	51.8	178	7 ADB46048	CanC fusi
7	305.5	29.3	124	5 AAU99739	Pyrodicti
8	305.5	29.3	124	7 ADB46052	CanE fusi
9	304.5	29.2	130	5 AAU99738	Pyrodicti
10	304.5	29.2	130	7 ADB46050	CanD fusi
11	96	9.2	255	7 ADE45056	CAMP fact
12	92	8.8	322	6 ABU33304	Protein e
13	91	8.7	255	5 ABP30785	Streptoco
14	87.5	8.4	978	7 ADC51492	Bacterial
15	86.5	8.3	284	5 ABB57211	Mouse isc
16	86.5	8.3	418	3 AAY77286	Streptomy
17	86.5	8.3	418	3 AAY78838	Amino aci
18	86.5	8.3	846	5 AAU82974	S. cerevi
19	86.5	8.3	846	6 ABR53317	Protein s
20	85.5	8.2	416	5 ABP65732	Bifidobac
21	85	8.2	338	5 ABB93924	Herbicida
22	83	8.0	626	2 AAU96010	Adenovir
23	83	8.0	626	6 ABU08321	M. bovis
24	83	8.0	675	5 ABP65592	Bifidobac
25	83	8.0	1246	5 AAU84295	Human end

ALIGNMENTS

RESULT 1

AAU99735
ID AAU99735 standard; protein; 207 AA.

XX AAU99735;

DT 07-OCT-2002 (first entry)

XX Pyrodictium abyssi Canulle A (CanA) protein.

XX Polymer; self-assembly; divalent cation; drug delivery; human body;
XX animal body; fibre; polymeric separation agent; coating composition;
XX biochip; nanomechanical component; optical switch; optical wave guide;
XX Canulle A; CanA.

OS Pyrodictium abyssi.

XX WO200244336-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US045001.

XX 30-NOV-2000; 2000US-0250426P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EU, Laiferty WM, Barton N, Chow K;

XX WPI; 2002-557536/59.

XX N-PSDB; ABK88271.

XX Producing a polypeptide polymer by self-assembly for use in lubricants
XX and coating compositions, comprises polymerizing polypeptides capable of
XX self-assembly in the presence of a divalent cation and template molecule.

XX Claim 9; Page 178-179; 182pp; English.

XX The present invention relates to a new method of producing a polypeptide
XX polymer by self-assembly. The method involves providing a number of
XX polypeptides capable of self-assembly in the presence of a divalent
XX cation and polymerising the polypeptides in the presence of a divalent
XX cation and a template molecule. The invention is useful for delivering a
XX drug to a location in the human or animal body. Polypeptides are useful
XX for encapsulating a molecule. The polymeric separation agent is useful
XX for isolating a chiral compound from a mixture. A nucleic acid is useful
XX for comparing a first sequence to a second sequence, where the first
XX sequence is a nucleic acid, and for identifying a feature in a particular

Abb90742 Human Tum
Aao17366 Human nid
Abu34449 Human tum
Abp27517 Streptoco
Aag51600 Arabidops
Aag06413 Arabidops
Aag51599 Arabidops
Aag06412 Arabidops
Aae36342 Arabidops
Abu24247 Protein e
Aaw43448 Tobacco 1
Abb47321 Listeria
Abb47302 Listeria
Abp73966 Candida a
Abu50834 Helicobac
Abu51217 Helicobac
Aar05710 TRY40. 3/
Aar06478 TRY40. 3/
Aar43675 Single ch
Aar99645 Single ch

26 83 1247 5 ABB90742
27 83 1247 5 AAO17366
28 83 1247 6 ABU34449
29 83 1370 5 ABP27517
30 82.5 7.9 238 3 AAG51600
31 82.5 7.9 238 3 AAG06413
32 82.5 7.9 243 3 AAG51599
33 82.5 7.9 243 3 AAG06412
34 82.5 7.9 296 6 AAE36342
35 82.5 7.9 296 6 ABU24247
36 81.5 7.8 557 2 AAW43448
37 81.5 7.8 862 5 ABB47321
38 81.5 7.8 2044 5 ABB47302
39 81 7.8 751 5 ABP73966
40 80.5 7.7 148 5 ABU50834
41 80.5 7.7 183 5 ABU51217
42 80.5 7.7 225 2 AAR05710
43 80.5 7.7 225 2 AAR06478
44 80.5 7.7 225 2 AAR43675
45 80.5 7.7 225 2 AAR99645

CC sequence. The polypeptide of the invention is useful in fibres, polymeric
 CC separation agents, coating compositions, biochips, nanomechanical
 CC components, optical switches and optical wave guides. The present amino
 CC acid sequence represents the Pyrodictium abyssii Cannule A (CanA) protein
 CC of the invention
 XX
 XX

SQ Sequence 207 AA;

Query Match 100.0%; Score 1042; DB 5; Length 207;
 Best Local Similarity 100.0%; Pred. No. 9.8e-95;
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTTLAAGIIASAAALALLAGFATTQSPNSFYATGTAAQVSEPIDVESHLSITPAA 60
 DB 1 VKYTTLAAGIIASAAALALLAGFATTQSPNSFYATGTAAQVSEPIDVESHLSITPAA 60
 QY 61 GAQSSDDIGVAIVWIKQVNDVKLVTLRNAEQLKPYFKYLIQITSGYETNSTALGNFS 120
 DB 61 GAQSSDDIGVAIVWIKQVNDVKLVTLRNAEQLKPYFKYLIQITSGYETNSTALGNFS 120
 QY 121 ETKAVISLDNPSAVIVLDKEDIAVLVDPKGTGYTNTSIWVGPEDKIIIVNETKPVAILNF 180
 DB 121 ETKAVISLDNPSAVIVLDKEDIAVLVDPKGTGYTNTSIWVGPEDKIIIVNETKPVAILNF 180
 QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207
 DB 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207

RESULT 2

ID ADB46044
 ID ADB46044 standard; protein; 207 AA.

AC ADB46044;

DT 04-DEC-2003 (first entry)

DE CanA fusion protein.

XX chimeric cannulae polypeptide; metabolite exchange; signal compound;
 KW CanA.

OS Pyrodictium abyssii.

XX WO2003070961-A1.

XX 28-AUG-2003.

XX 18-FEB-2003; 2003WO-US004828.

XX 15-FEB-2002; 2002US-0357406P.

XX (DIVE-) DIVERSA CORP.

XX Barton NR, O'donoghue E, Frey G;

XX WPI; 2003-689787/65.

XX N-PSDB; ADB46043.

XX New chimeric polypeptide comprising a first domain comprising a cannulae
 PT polypeptide and a second domain comprising a heterologous polypeptide or
 PT peptide useful for selecting and purifying chiral compositions from
 PT racemic mixtures.

PS Claim 3; Page 81-82; 89pp; English.

XX The present invention relates to a chimeric cannulae polypeptide and a
 CC second domain having a heterologous polypeptide or peptide. The
 CC polypeptides, nanotubule and methods are useful for identifying,
 CC separating and synthesizing proteins or ligands, and for selecting
 CC purifying chiral compositions from racemic mixtures. The nanotubule is
 CC useful for cells to exchange metabolites, genetic information or signal
 CC compounds. The present sequence represents CanA fusion protein.

XX SQ Sequence 207 AA;

Query Match 100.0%; Score 1042; DB 7; Length 207;
 Best Local Similarity 100.0%; Pred. No. 9.8e-95;
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTTLAAGIIASAAALALLAGFATTQSPNSFYATGTAAQVSEPIDVESHLSITPAA 60
 DB 1 VKYTTLAAGIIASAAALALLAGFATTQSPNSFYATGTAAQVSEPIDVESHLSITPAA 60
 QY 61 GAQSSDDIGVAIVWIKQVNDVKLVTLRNAEQLKPYFKYLIQITSGYETNSTALGNFS 120
 DB 61 GAQSSDDIGVAIVWIKQVNDVKLVTLRNAEQLKPYFKYLIQITSGYETNSTALGNFS 120
 QY 121 ETKAVISLDNPSAVIVLDKEDIAVLVDPKGTGYTNTSIWVGPEDKIIIVNETKPVAILNF 180
 DB 121 ETKAVISLDNPSAVIVLDKEDIAVLVDPKGTGYTNTSIWVGPEDKIIIVNETKPVAILNF 180
 QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207
 DB 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207

RESULT 3

AAU99736

ID AAU99736 standard; protein; 170 AA.

XX AAU99736;

DT 07-OCT-2002 (first entry)

DE Pyrodictium abyssii Cannule B (CanB) protein.

XX Polymer; self-assembly; divalent cation; drug delivery; human body;
 KW animal body; fibre; polymeric separation agent; coating composition;
 KW biochip; nanomechanical component; optical switch; optical wave guide;
 KW Cannule B; CanB.

XX Pyrodictium abyssii.

XX WO200244336-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US045001.

XX 30-NOV-2000; 2000US-0250426P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EJ, Lafferty WM, Barton N, Chow K;

XX WPI; 2002-557536/59.

XX N-PSDB; ABK88272.

XX Producing a polypeptide polymer by self-assembly for use in lubricants
 PT and coating compositions, comprises polymerizing polypeptides capable of
 PT self-assembly in the presence of a divalent cation and template molecule.

XX Claim 9; Page 179-180; 182pp; English.

XX The present invention relates to a new method of producing a polypeptide
 CC polymer by self-assembly. The method involves providing a number of
 CC polypeptides capable of self-assembly in the presence of a divalent
 CC cation and polymerising the polypeptides in the presence of a divalent
 CC cation and a template molecule. The invention is useful for delivering a
 CC drug to a location in the human or animal body. Polypeptides are useful
 CC for encapsulating a molecule. The polymeric separation agent is useful
 CC for isolating a chiral compound from a mixture. A nucleic acid is useful
 CC for comparing a first sequence to a second sequence, where the first
 CC sequence is a nucleic acid, and for identifying a feature in a particular
 CC sequence. The polypeptide of the invention is useful in fibres, polymeric

CC separation agents, coating compositions, biochips, nanomechanical
CC components, optical switches and optical wave guides. The present amino
CC acid sequence represents the Pyrodictium abyssii Cannule B (CancB) protein
CC of the invention
XX
SQ Sequence 170 AA;
Query Match 53.1%; Score 553.5; DB 5; Length 170;
Best Local Similarity 60.2%; Pred. No. 1.7e-46;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;
QY 1 VKYTTLAAGIIASAAALALLAGFATTQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
DB 1 VKPTALALAGIIASADLALLAGFATTQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
QY 61 GAQSDDIGYAIWIKQVNDVKLVLRNAEQPKYFKYLQIQITSGYETNSTALGNFS 120
DB 61 GAQSQDIGYFNVTAKQVNVTKIKVTLANAEOQLKPYFKYLQIVLKSEVA-----D 111
QY 121 ETKAVISLDNPSAVIVLDKEDIAVLVDPKGTGTYNTSIWVGPEDPKIIVYNETKPVAILNF 180
DB 112 EIRKAVISIDKPSAVIILDSQDF-----DSNNRAKISA 143
QY 181 KAFYEAKGMLFDSLPIFNQVLOV 206
DB 144 TAYEAKGMLFDSLPIFNQVLSV 169
RESULT 4
ADB46046
ID ADB46046 standard; protein; 170 AA.
AC ADB46046;
DT 04-DEC-2003 (first entry)
XX CanB fusion protein.
XX chimeric cannulae polypeptide; metabolite exchange; signal compound;
XX CanB.
XX Pyrodictium abyssii.
XX WO2003070961-A1.
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004828.
XX 15-FEB-2002; 2002US-0357406P.
XX (DIVE-) DIVERSA CORP.
PI Barton NR, O'donoghue E, Frey G;
XX WPI; 2003-689787/65.
DR N-PSDB; ADB46045.
XX New chimeric polypeptide comprising a first domain comprising a cannulae
PT polypeptide and a second domain comprising a heterologous polypeptide
PT peptide useful for selecting and purifying chiral compositions from
XX racemic mixtures.
XX Claim 3; Page 82-83; 89pp; English.
XX The present invention relates to a chimeric cannulae polypeptide and a
CC second domain having a heterologous polypeptide or peptide. The
CC polypeptides, nanotubule and methods are useful for identifying,
CC separating and synthesizing proteins or ligands, and for selecting
CC purifying chiral compositions from racemic mixtures. The nanotubule is
CC useful for calls to exchange metabolites, genetic information or signal
CC compounds. The present sequence represents CanB fusion protein.
XX

SQ Sequence 170 AA;
Query Match 53.1%; Score 553.5; DB 7; Length 170;
Best Local Similarity 60.2%; Pred. No. 1.7e-46;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;
QY 1 VKYTTLAAGIIASAAALALLAGFATTQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
DB 1 VKPTALALAGIIASADLALLAGFATTQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
QY 61 GAQSDDIGYAIWIKQVNDVKLVLRNAEQPKYFKYLQIQITSGYETNSTALGNFS 120
DB 61 GAQSQDIGYFNVTAKQVNVTKIKVTLANAEOQLKPYFKYLQIVLKSEVA-----D 111
QY 121 ETKAVISLDNPSAVIVLDKEDIAVLVDPKGTGTYNTSIWVGPEDPKIIVYNETKPVAILNF 180
DB 112 EIRKAVISIDKPSAVIILDSQDF-----DSNNRAKISA 143
QY 181 KAFYEAKGMLFDSLPIFNQVLOV 206
DB 144 TAYEAKGMLFDSLPIFNQVLSV 169
RESULT 5
AAU99737
ID AAU99737 standard; protein; 178 AA.
XX AAU99737;
AC AAU99737;
DT 07-OCT-2002 (first entry)
XX Pyrodictium abyssii Cannule C (CancC) protein.
XX Polymer; self-assembly; divalent cation; drug delivery; human body;
KW animal body; fibre; polymeric separation agent; coating composition;
KW biochip; nanomechanical component; optical switch; optical wave guide;
KW Cannule C; CancC.
XX Pyrodictium abyssii.
XX WO200244336-A2.
XX 06-JUN-2002.
XX 30-NOV-2001; 2001WO-US045001.
XX 30-NOV-2000; 2000US-0250426P.
PR (DIVE-) DIVERSA CORP.
XX Short J, Mathur EJ, Lafferty WM, Barton N, Chow K;
XX WPI; 2002-557536/59.
DR N-PSDB; ABK88273.
XX Producing a polypeptide polymer by self-assembly for use in lubricants
PT and coating compositions, comprises polymerizing polypeptides capable of
PT self-assembly in the presence of a divalent cation and template molecule.
XX Claim 9; Page 180; 182pp; English.
XX The present invention relates to a new method of producing a polypeptide
CC polymer by self-assembly. The method involves providing a number of
CC polypeptides capable of self-assembly in the presence of a divalent
CC cation and polymerising the polypeptides in the presence of a divalent
CC cation and a template molecule. The invention is useful for delivering a
CC drug to a location in the human or animal body. Polypeptides are useful
CC for encapsulating a molecule. The polymeric separation agent is useful
CC for isolating a chiral compound from a mixture. A nucleic acid is useful
CC for comparing a first sequence to a second sequence, where the first
CC sequence is a nucleic acid, and for identifying a feature in a particular
CC sequence. The polypeptide of the invention is useful in fibres, polymeric
CC separation agents, coating compositions, biochips, nanomechanical
CC

CC acid sequence represents the Pyrodicticum abyssii Cannule E (CanE) partial
 XX protein of the invention
 SQ Sequence 124 AA;

Query Match 29.3%; Score 305.5; DB 5; Length 124;
 Best Local Similarity 48.8%; Pred. No. 3.8e-22;
 Matches 78; Conservative 14; Mismatches 31; Indels 37; Gaps 5;

QY 33 SFYATGTAQAVSEPIDVESHG-SITPAAGAGSDDIGYAIWIKQVNDVKLVILRNA 91
 Db 1 SFYATGTAATSEPIDVSNLNTAIPAAAGQSGVIGSTTIENKTDVNVVKLKITLANA 60
 QY 92 EQLKPYFKYLIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIAVLDPDKTG 151
 Db 61 EQLKPYFDYLIQIVLKS-VDSN-----EIKAVLSLEKPSAVIILDNEF-----QG 104
 QY 152 YNTISWVPGEPDKIIVYNETKPVAILNFKAFYEAKGML 191
 Db 105 GDNQ-----COIDATAYYEAKGML 124

RESULT 8

AD846052

ID ADB46052 standard; protein; 124 AA.

XX AC ADB46052;

DT 04-DEC-2003 (first entry)

XX CanE fusion protein.

XX chimeric cannulae polypeptide; metabolite exchange; signal compound;
 KW CanE.

XX Pyrodicticum abyssii.

XX WO2003070961-A1.

XX 28-AUG-2003.

XX 18-FEB-2003; 2003WO-US004828.

XX 15-FEB-2002; 2002US-0357406P.

XX (DIVE-) DIVERSA CORP.

PI Barton NR, O'donoghue E, Frey G;

DR WPI; 2003-689787/65.

DR N-PSDB; ADB46051.

XX New chimeric polypeptide comprising a first domain comprising a cannulae
 PT polypeptide and a second domain comprising a heterologous polypeptide or
 PT peptide useful for selecting and purifying chiral compositions from
 PT racemic mixtures.

XX Disclosure; Page 85; 89pp; English.

XX The present invention relates to a chimeric cannulae polypeptide and a
 CC second domain having a heterologous polypeptide or peptide. The
 CC polypeptides, nanotubule and methods are useful for identifying,
 CC separating and synthesizing proteins or ligands, and for selecting and
 CC purifying chiral compositions from racemic mixtures. The nanotubule is
 CC useful for cells to exchange metabolites, genetic information or signal
 CC compounds. The present sequence represents CanE fusion protein.

XX Sequence 124 AA;

Query Match 29.3%; Score 305.5; DB 7; Length 124;
 Best Local Similarity 48.8%; Pred. No. 3.8e-22;
 Matches 78; Conservative 14; Mismatches 31; Indels 37; Gaps 5;

QY 33 SFYATGTAQAVSEPIDVESHG-SITPAAGAGSDDIGYAIWIKQVNDVKLVILRNA 91
 Db 1 SFYATGTAATSEPIDVSNLNTAIPAAAGQSGVIGSTTIENKTDVNVVKLKITLANA 60
 QY 92 EQLKPYFKYLIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIAVLDPDKTG 151
 Db 61 EQLKPYFDYLIQIVLKS-VDSN-----EIKAVLSLEKPSAVIILDNEF-----QG 104
 QY 152 YNTISWVPGEPDKIIVYNETKPVAILNFKAFYEAKGML 191
 Db 105 GDNQ-----COIDATAYYEAKGML 124

RESULT 9

AAU99738

ID AAU99738 standard; protein; 130 AA.

XX AC AAU99738;

DT 07-OCT-2002 (first entry)

XX Pyrodicticum abyssii Cannule D (CanD) partial protein.

XX Polymer; self-assembly; divalent cation; drug delivery; human body;
 KW animal body; fibre; polymeric separation agent; coating composition;
 KW biochip; nanomechanical component; optical switch; optical wave guide;
 KW Cannule D; CanD partial.

XX Pyrodicticum abyssii.

XX WO200244336-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US045001.

XX 30-NOV-2000; 2000US-0250426P.

XX (DIVE-) DIVERSA CORP.

PI Short J, Mathur EJ, Lafferty WM, Barton N, Chow K;

DR WPI; 2002-557536/59.

DR N-PSDB; ABX88274.

XX Producing a polypeptide polymer by self-assembly for use in lubricants
 PT and coating compositions, comprises polymerizing polypeptides capable of
 PT self-assembly in the presence of a divalent cation and template molecule.

XX Claim 9; Page 181; 182pp; English.

XX The present invention relates to a new method of producing a polypeptide
 CC polymer by self-assembly. The method involves providing a number of
 CC polypeptides capable of self-assembly in the presence of a divalent
 CC cation and polymerising the polypeptides in the presence of a divalent
 CC cation and a template molecule. The invention is useful for delivering a
 CC drug to a location in the human or animal body. Polypeptides are useful
 CC for encapsulating a molecule. The polymeric separation agent is useful
 CC for isolating a chiral compound from a mixture. A nucleic acid is useful
 CC for comparing a first sequence to a second sequence, where the first
 CC sequence is a nucleic acid, and for identifying a feature in a particular
 CC sequence. The polypeptide of the invention is useful in fibres, polymeric
 CC separation agents, coating compositions, biochips, nanomechanical
 CC components, optical switches and optical wave guides. The present amino
 CC acid sequence represents the Pyrodicticum abyssii Cannule D (CanD) partial
 CC protein of the invention. Note: The specification states that this
 CC sequence is encoded by the Pyrodicticum abyssii Cannule D (CanD) partial
 CC gene (ABX88274) but this sequence only encodes part of the protein
 CC (residues 1-103)

XX Sequence 130 AA;

Query Match 29.2%; Score 304.5; DB 5; Length 130;

Best Local Similarity 47.2%; Pred. No. 5.1e-22;
Matches 75; Conservative 22; Mismatches 33; Indels 29; Gaps 6;

QY 33 SFYATGTAQAVSEPIDVESHGSLTPAAGQSDDIGYAIWIKQNDVKLVKTLRNAE 92
Db 1 SFYATGTAQAVSEPIDVESHGSLTPAAGQSDDIGYAIWIKQNDVKLVKTLRNAE 60

QY 93 QLKPYFYKYLQITSGVETNLTALGNFSETKAVISLNDPSAVIVLDKEDIAVLVLPDKTGY 152
Db 61 QLRPFYKYLIIKLVS-LDSN---GNESEKGMITLWKPVAIVILDHED-----F 105

QY 153 TNSIWPBPDKIIIVNTEKPVAILNFKAFYEAKGML 191
Db 106 NND---IDGD-----NQCC-----IDATAYYEAKGML 130

RESULT 10
ADB46050
ID ADB46050 standard; protein; 130 AA.
XX
AC ADB46050;
XX
DT 04-DEC-2003 (first entry)
XX
DE CanD fusion protein.
XX
KW chimeric cannulae polypeptide; metabolite exchange; signal compound;
KW CanD.
XX
OS Pyrodictium abyssi.
XX
PN W02003070961-A1.
XX
PD 28-AUG-2003.
XX
PF 18-FEB-2003; 2003WO-US004828.
XX
PR 15-FEB-2002; 2002US-0357406P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Barton NR, O'donoghue E, Frey G;
XX
DR WPI; 2003-689787/65.
DR N-PSDB; ADB46049.
XX
New chimeric polypeptide comprising a first domain comprising a cannulae polypeptide and a second domain comprising a heterologous polypeptide or peptide useful for selecting and purifying chiral compositions from racemic mixtures.
XX
PS Disclosure; Page 84; 89pp; English.
XX
CC The present invention relates to a chimeric cannulae polypeptide and a second domain having a heterologous polypeptide or peptide. The polypeptides, nanotubule and methods are useful for identifying, separating and synthesizing proteins or ligands, and for selecting and purifying chiral compositions from racemic mixtures. The nanotubule is useful for cells to exchange metabolites, genetic information or signal compounds. The present sequence represents CanD fusion protein.
XX
SQ Sequence 130 AA;

Query Match 29.2%; Score 304.5; DB 7; Length 130;
Best Local Similarity 47.2%; Pred. No. 5.1e-22;
Matches 75; Conservative 22; Mismatches 33; Indels 29; Gaps 6;

QY 33 SFYATGTAQAVSEPIDVESHGSLTPAAGQSDDIGYAIWIKQNDVKLVKTLRNAE 92
Db 1 SFYATGTAQAVSEPIDVESHGSLTPAAGQSDDIGYAIWIKQNDVKLVKTLRNAE 60

QY 93 QLKPYFYKYLQITSGVETNLTALGNFSETKAVISLNDPSAVIVLDKEDIAVLVLPDKTGY 152
Db 61 QLRPFYKYLIIKLVS-LDSN---GNESEKGMITLWKPVAIVILDHED-----F 105

QY 153 TNSIWPBPDKIIIVNTEKPVAILNFKAFYEAKGML 191
Db 106 NND---IDGD-----NQCC-----IDATAYYEAKGML 130

RESULT 11
ADB45056
ID ADB45056 standard; protein; 255 AA.
XX
AC ADB45056;
XX
DT 29-JAN-2004 (first entry)
XX
DE CAMP factor, SEQ ID 4.
XX
KW Immunogenic; cyclic adenosine monophosphate factor; CAMP factor; mastitis; listeriosis; antibacterial; antiinflammatory; vaccine; immunostimulant; gene therapy.
KW Streptococcus agalactiae.
XX
OS
FH Key Location/Qualifiers
FT Peptide 1..29 /note= "Signal peptide"
FT Protein 30..255 /note= "Mature protein"
FT
XX
PN W02003091437-A1.
XX
PD 06-NOV-2003.
XX
PF 24-APR-2003; 2003WO-CA000587.
XX
PR 26-APR-2002; 2002US-00134021.
XX
PA (UYSA-) UNIV SASKATCHEWAN.
XX
PI Potter AA, Perez-Casal J, Fontaine M, Song X;
XX
DR WPI; 2003-877332/81.
DR N-PSDB; ADE45055.
XX
New cyclic adenosine monophosphate(CAMP)-3 polypeptide comprising streptococcal CAMP factor epitopes, useful in treating or preventing bacterial infection or as vaccine or diagnostic reagent.
XX
PS Claim 7; Fig 3; 73pp; English.
XX
CC The present invention relates to immunogenic proteins comprising one or more cyclic adenosine monophosphate (CAMP) factor epitopes from more than one bacterial species. The proteins and CAMP factors are useful in detecting Streptococcus antibodies in a biological sample, eliciting an immunological response against streptococcal infections, used in vaccine compositions or as diagnostic reagents or in preventing or treating streptococcal infection which causes mastitis. The proteins are also useful in preventing or treating listeriosis. The present sequence is a CAMP factor.
XX
SQ Sequence 255 AA;

Query Match 9.2%; Score 96; DB 7; Length 255;
Best Local Similarity 25.0%; Pred. No. 0.59;
Matches 53; Conservative 31; Mismatches 84; Indels 44; Gaps 10;

QY 5 TLAAGIIASAALALAGFATTCSPNSFYATGTAQAVSEPIDVES-HLGSIITPAAGAQ 63
Db 12 TLVAGALLFSPAVLEHVADQVTTTPQVNVHNSNNQAQAKLDQDSIQLRNKID--NVQ 69

QY 64 GSD---DIGYAIWIKQNDVKLVKTLR-----NARQLKPYFYKYLQIT 106
Db 70 GTDYKPVNEAITSVE-----KLKTSLRANPETVYDLNSIGSRVEALTDVIE--AITFS 121

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins

XX Sequence 255 AA;
 Query Match 8.7%; Score 91; DB 5; Length 255;
 Best Local Similarity 24.5%; Pred. NO. 1.8;
 Matches 52; Conservative 31; Mismatches 85; Indels 44; Gaps 10;

QY 5 TLALAGIIASAAALLAGFATTQSPINSPFATCTAQAQVSEPIDVES-HLGSITPAAGAQ 63
 DB 12 TLVAGALLFPSPVLEHADVDTTPOVNVHNSNNOAQOQMAQKLDODSIOLENIKD--NVQ 69
 QY 64 GSD---DIGYAIWIKQVNDVKLKVTLR-----NAEQKPYFKYLQIQIT 106
 DB 70 GTDYKTVNEAITSVE-----KLKTSURANPETVYDLNSIGSRVEALTDVIE--AITFS 121
 QY 107 SGYETNSTALGN---PSETKAVISLNDPSAVI-----VLDKEDIAVLYPKDTGYT 153
 DB 122 TOHLANKVSOANIDMGFGITKLIVRIIDPFASVDSIKAQVNDVKALEOKVLTYPDLKPTD 181
 QY 154 NTSIWVPGEPPKIIIVNE--TKPVAILNFKAF 183
 DB 182 RATIYTKSKUDKEI-WNTRFRDRKKVLNVKEF 212

RESULT 14
 ADCS1492
 ID ADCS1492 standard; protein; 978 AA.
 XX AC ADCS1492;
 XX AC ADCS1492;

DT 18-DEC-2003 (first entry)

XX Bacterial beta-hexosaminidase gene SEQ ID NO:8.

XX beta-hexosaminidase; beta-glucosidase; sphingo-glycolipid.

XX Paenibacillus sp.

OS JP2003061663-A.

XX 04-MAR-2003.

XX 21-AUG-2001; 2001JP-00249782.

XX 21-AUG-2001; 2001JP-00249782.

XX (SANG-) SANGAKU RENKEI KIKO KYUSHU KK.

XX WPI; 2003-691955/66.

DR N-PSDB; ADCS1491.

XX Novel exo-type ganglioside degradation enzymes e.g. beta- glucosidase and
 PT beta-hexosaminidase, useful for producing sphingoglycolipids, are derived
 PT from Paenibacillus species of strain TS12 FERM P-18416.

XX Claim 4; SEQ ID NO 8; 23pp; Japanese.

XX The invention relates to a novel recombinant polypeptide, encoded by beta
 CC - hexosaminidase gene derived from Paenibacillus sp. or encoded by beta-
 CC glucosidase gene derived from Escherichia coli. A gene of the invention

CC is useful for producing sphingo- glycolipids. The present sequence is
 CC used in the exemplification of the invention.

XX Sequence 978 AA;

Query Match 8.4%; Score 87.5; DB 7; Length 978;
 Best Local Similarity 25.6%; Pred. No. 28;
 Matches 43; Conservative 17; Mismatches 69; Indels 39; Gaps 6;
 QY 48 DVESHLSGITPAAGAGSDDDIGYAIWIKQVNDVKLKVTLRANAEQLKPYFKYLQIQITS 107
 DB 477 DFKERLIEHTPRLQNGKIKFFADPIVW-----ELPIVQINSEWKMDEGTGVVWVSDTS 528
 QY 108 GYETNSTALGNFSETKAV---ISLDNPSAVIVLDKEDIAVLYPKDTGYTNTSIWVCEP 163
 DB 529 GY-LNGTLVGGAKWTAGKQNGVSGDSSGVINLGGQDI-----IGNWTAAVVWVVGQP 580
 QY 164 D-----KLIIVNETKPVAILNFKAFYAEKGMFLPDSLP 196
 DB 581 NNTNNTLLSGTTSIAIKINQNTKTKVGI---TIYGTQDYTYNYSIP 624

RESULT 15

ABB57211

ID ABB57211 standard; protein; 284 AA.

XX AC ABB57211;

XX 07-MAR-2002 (first entry)

XX Mouse ischaemic condition related protein sequence SEQ ID NO:514.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 XX vasospastic ischaemia; ischaemic condition; ischaemic disease.

OS Mus musculus.

PN WO200188188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001WO-JP004192.

XX 18-MAY-2000; 2000JP-00145977.

XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX WPI; 2002-034733/04.

DR N-PSDB; ABI99523.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 PT genes.

PS Claim 2; Page 1418-1419; 2690pp; English.

XX The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (I) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (I). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
 CC protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles produced by these genes are used as an
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention

```
XX
SQ      Sequence 284 AA;
      Query Match      8.3%; Score 86.5; DB 5; Length 284;
      Best Local Similarity 26.2%; Pred. No. 6;
      Matches 45; Conservative 28; Mismatches 68; Indels 31; Gaps 8;

QY      1 VKYTTLAIAGIIASAALALLAGFATTQSPNLSFYATGT---AQAYSEPIDVESHLSIT 57
Db      |.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
      126 VQATLAVANITNADSATRELLA-----QTLRN--ALGINKLSQILSDREEIAHHMQSTL 178
QY      58 PAAGAQSDDIGYAIWMIKDQVNDVKLVTLRMAEOLKPYFKYLQIQTSGYETNSTALG 117
Db      |.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
      179 DDA---TDDWGIKVERV--EIKDVKLPVQLQRA-----MAAEAEAREARAKVIA 223
QY      118 NFSEYKAVISLDNPSAVIVLDKEDIAVLYPDKTGYTNTSIWVPGEPDKLIIVY 169
Db      |.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
      224 AEGEMNASRALKKESMWIT---ESPAAL---QLRYLQTLTTIAAEKNSTIVF 269
```

Search completed: April 6, 2004, 12:38:53
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 11:43:46 ; Search time 22 Seconds
(without alignments)
485.753 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTTLATAGIIASAAALAL.....EGMLFDSLPIVFNQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2.6/prodata/2/iaa/5A COMB.pcp.*
2: /cgn2.6/prodata/2/iaa/5B COMB.pcp.*
3: /cgn2.6/prodata/2/iaa/6A COMB.pcp.*
4: /cgn2.6/prodata/2/iaa/6B COMB.pcp.*
5: /cgn2.6/prodata/2/iaa/PCUS COMB.pcp.*
6: /cgn2.6/prodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	10.1	651	4	US-09-543-681A-8284
2	94	9.0	208	4	US-09-252-991A-32467
3	83	8.0	626	4	US-09-485-717-2
4	80.5	7.7	225	6	5455030-5
5	80.5	7.7	256	4	US-09-107-532A-4072
6	79.5	7.6	769	3	US-09-320-878-12
7	79.5	7.6	769	4	US-09-141-908-10
8	79.5	7.6	769	4	US-09-657-440-12
9	79.5	7.6	809	3	US-09-105-537-24
10	79.5	7.6	1461	4	US-09-976-594-531
11	79.5	7.6	3782	3	US-09-105-537-4
12	79	7.6	1180	5	US-09-224-024-28
13	79	7.6	1180	5	US-09-489-039A-13379
14	78.5	7.5	355	4	US-09-489-039A-13382
15	78	7.5	408	4	US-09-489-039A-13379
16	78	7.5	455	4	US-09-711-164-416
17	77	7.4	555	4	US-09-489-039A-13382
18	77	7.4	927	4	US-09-841-786-3
19	77	7.4	3241	4	US-09-841-786-1
20	76.5	7.3	225	4	US-09-543-681A-4987
21	76.5	7.3	383	3	US-09-413-814-12
22	76.5	7.3	616	1	US-08-385-370-2
23	76.5	7.3	616	1	US-08-385-370-4
24	76.5	7.3	655	1	US-08-469-202-27
25	76.5	7.3	655	2	US-08-484-434C-34
26	76.5	7.3	655	4	US-09-384-361-34
27	76.5	7.3	672	4	US-09-489-039A-12101

28 76 7.3 669 4 US-09-071-035-264 Sequence 264, App
29 76 7.3 1026 1 US-08-194-290-7 Sequence 7, Appli
30 76 7.3 1026 3 US-08-614-377A-7 Sequence 7, Appli
31 76 7.3 1026 3 US-09-142-648B-7 Sequence 7, Appli
32 76 7.3 1638 4 US-09-071-035-258 Sequence 258, App
33 76 7.3 1638 4 US-09-071-035-262 Sequence 262, App
34 76 7.3 1638 4 US-09-071-035-266 Sequence 266, App
35 76 7.3 1747 4 US-09-134-000C-5999 Sequence 5999, App
36 75.5 7.2 274 4 US-09-813-659-30 Sequence 30, Appl
37 75.5 7.2 274 4 US-09-549-067A-30 Sequence 30, Appl
38 75.5 7.2 291 4 US-09-543-681A-6165 Sequence 6165, Ap
39 75.5 7.2 302 1 US-08-121-054C-30 Sequence 30, Appl
40 75.5 7.2 302 1 US-08-539-436-30 Sequence 30, Appl
41 75.5 7.2 302 4 US-09-813-659-32 Sequence 32, Appl
42 75.5 7.2 302 4 US-09-549-067A-32 Sequence 32, Appl
43 75.5 7.2 399 4 US-09-252-991A-20610 Sequence 20610, A
44 75.5 7.2 728 4 US-09-252-991A-28459 Sequence 28459, A
45 75.5 7.2 1251 4 US-09-252-991A-17263 Sequence 17263, A

ALIGNMENTS

RESULT 1

US-09-543-681A-8284
; Sequence 8284, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709, 1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8284
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8284

Query Match 10.1%; Score 105; DB 4; Length 651;
Best Local Similarity 22.0%; Pred. No. 0.013; 89; Indels 54; Gaps 7;
Matches 48; Conservative 27; Mismatches 21
QY 14 SAAALALLAGFATQSPPLNSFYATCTAQAVSEPIDVESHLSIT-PAAGAQSDDIGYAI 72
Db 130 SRHVLILIDGVTNQAGISGSDMSQL-----FLSLVQRIEYIRGSRNAVYGSDAIGGVI 184
QY 73 VIKDQNDVKLVTLRANAEOLKPKYKL-QIQTSGYETNSTALGNFSEKAVISLNDP 131
Db 185 NFITKRPND---GLTNAGIGSHGYQNYNGSIQOKVGERTTLSAAGGYTYTKGIHATD-- 239
QY 132 SAVIVLKDIEDIAVLDPDKTGYNTSIW-----VPE 162
Db 240 -----DSVPVAPPHDRHGFNMKSYWFGIDHFNDFQGMKGYGYNRTTYVASYPGN 293
QY 163 PDKIIVNETKPVAILNFKAFYEAKGMLFDLSLPVFN 200
Db 294 TDQSKTYNR-----NVEAGIKFSGDSYFSLITVYS 324

RESULT 2

US-09-252-991A-32467
; Sequence 32467, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

DB 180 SELVENQPDVLPVKRDSLTLSDLFGMTNQDNKIVVKNATK 220

RESULT 4
5455030-5
; Patent No. 5455030
; APPLICANT: LAMNER, ROBERT C.; BIRD, ROBERT E.; HARDMAN, KARL
; TITLE OF INVENTION: IMMUNOTHERAPY USING SINGLE CHAIN
; POLYPEPTIDE BINDING MOLECULES
; NUMBER OF SEQUENCES: 24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/40,440
; FILING DATE: 1-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 512,910
; FILING DATE: 25-APR-1990
; APPLICATION NUMBER: 299,617
; FILING DATE: 19-JAN-1989
; APPLICATION NUMBER: 92,110
; FILING DATE: 02-SEP-1987
; APPLICATION NUMBER: 902,971
; FILING DATE: 01-SEP-1986
; SEQ ID NO:5
; LENGTH: 225
; 5455030-5

Query Match 7.7%; Score 80.5; DB 6; Length 225;
Best Local Similarity 24.5%; Pred. No. 1-2;
Matches 40; Conservative 22; Mismatches 66; Indels 35; Gaps 8;

QY 22 AGFATTQSLNFGYATGTAAVSEPIDVESHLGSIPTAAGACGGDDIGYAIWIKDQVND 81
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 26 ASSSVSSYLHWFFQQKSIAKAFKNGGLVKPGSLKCSCAASGFTFISYGMSWVR-QTPD 84
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 82 VKLK-VILRNAEQLKPKYLIQITSGYETNSTALGNFSETKAVISLDNFSAIVLD- 138
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 85 KRLWEVATISSGS-----TYYTPDSVK-GRFT-----ISRDNAKDTLYLQM 125
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 139 ----KEDIAVLVPDK-----TGVTNTSIVWPGEPPDKIIYN 170
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 126 SGLKSEDTRMYCARRITTVLTVDIYANDYWGPGSP-KLWIYS 167
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 5
US-09-107-532A-4072
; Sequence 4072, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4072:
SEQUENCE CHARACTERISTICS:
LENGTH: 256 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...256
SEQUENCE DESCRIPTION: SEQ ID NO: 4072:
US-09-107-532A-4072

Query Match 7.7%; Score 80.5; DB 4; Length 256;
Best Local Similarity 25.7%; Pred. No. 1.5;
Matches 43; Conservative 24; Mismatches 61; Indels 39; Gaps 8;
QY 44 SEPIDVESHLSITPAAGAGSDDIGYAIWIKD--QVNDVKLVTLRNEAQLKPYFKYL 101
DB 91 SEQLDNQGH-----EKQGVNGAKFSYVDSDILQKMDVKDLTTDQIESQLKDRVKKL 143
QY 102 ---QIQTSGYEINS-----TALGNFS-----ETKAVISLDNPSANVILDKEDIAVLVP- 147
DB 144 SSDQLKSVNGEYTKTIDQQTGTGFEFSVEVQANQKQAYIIVNESSPENISNSEDILLTPV 203
QY 148 -DKTGYNTSIWV-----PGEPKIIVYNETKPVAILNFKAFYE 185
DB 204 SDRNGFLKDVWYIPKSEASQPKKEVKIVSTGVK-----KNFFE 243

RESULT 6
US-09-320-878-12
; Sequence 12, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-12

Query Match 7.6%; Score 79.5; DB 3; Length 769;

Best Local Similarity 19.8%; Pred. No. 11;
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;
QY 17 ALALLAGPATTQSPPLNSFYATGTAQAVSEP-IDVESHLSITPAAGAGSDDIGYAIWVI 75
DB 428 AVRATGGYATVQLGSHSTIEAGQVYGVKVSPLLKLTGTHKLTISGFAMSGATPLSLELGVW 487
QY 76 KQVNDVKLVTLRNEAQLKPYFKYLQIQTSGYEINSTAL-GNFSETKAVISLDNPSAV 134
DB 488 TPAADATTIAKAVESARKARTAVVFAYDDGTGTEGVRPNLSLPGTQDKLISAVADANPNTI 547
QY 135 IVDKEDIADVLP--DKTGYNTSIWVPG-----EPDKIIVYNETKPVAILNFKAFYEAK 188
DB 548 VVLTGSG-SVLMFWLSKTRAV-LDMWYPGQAGAEATAALLYGDVNP-----S 592
QY 189 GMLFDSLPIVFN 200
DB 593 GKLTQSFPAEN 604

RESULT 7
US-09-141-908-10
; Sequence 10, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PRQV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-10

Query Match 7.6%; Score 79.5; DB 4; Length 769;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;
QY 17 ALALLAGPATTQSPPLNSFYATGTAQAVSEP-IDVESHLSITPAAGAGSDDIGYAIWVI 75
DB 428 AVRATGGYATVQLGSHSTIEAGQVYGVKVSPLLKLTGTHKLTISGFAMSGATPLSLELGVW 487
QY 76 KQVNDVKLVTLRNEAQLKPYFKYLQIQTSGYEINSTAL-GNFSETKAVISLDNPSAV 134
DB 488 TPAADATTIAKAVESARKARTAVVFAYDDGTGTEGVRPNLSLPGTQDKLISAVADANPNTI 547
QY 135 IVDKEDIADVLP--DKTGYNTSIWVPG-----EPDKIIVYNETKPVAILNFKAFYEAK 188
DB 548 VVLTGSG-SVLMFWLSKTRAV-LDMWYPGQAGAEATAALLYGDVNP-----S 592
QY 189 GMLFDSLPIVFN 200
DB 593 GKLTQSFPAEN 604

RESULT 8


```

US-09-657-440-12
; Sequence 12, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062202120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 769
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-657-440-12

Query Match          7.6%; Score 79.5; DB 4; Length 769;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;

QY 17 ALALLAGFATTQSPFLNSFYATGTAQAYSEP-IDVESHLGSIITPAAGAGSDDIGYAIWVI 75
Db 428 AVRATGCGYATVQLGSHPTIEAGQVYGVKSSPLLKLTGKTHKLTISGFAVSATPLSLELGWV 487
QY 76 KDQVNDVKLVTLRNAPQLPYFYKYLQIQTSGVETNSTAL-GNFSKTKAVISLDNPSAV 134
Db 488 TPAAADATIAKAVESARKARTAVVAFYDDGTGVDREPNLSLPGTQDKLISAVADANPNTI 547
QY 135 IVDKEDIAVLYP--DKTGYTNTSIIWPG---EPDKIIVYNETKPVAILNFKAFYAKE 188
Db 548 VLNLTGS-SVLMPWLSKTRAV-LDMWYPGQAGAEATAALLYGVNP-----S 592
QY 189 GMLFDSLPIVFEN 200
Db 593 GKLTQSPFAAEN 604

RESULT 9
US-09-105-537-24
; Sequence 24, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 809
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-105-537-24

Query Match          7.6%; Score 79.5; DB 3; Length 809;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;

QY 17 ALALLAGFATTQSPFLNSFYATGTAQAYSEP-IDVESHLGSIITPAAGAGSDDIGYAIWVI 75

```

; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO: 4
; LENGTH: 3782
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-4

Query Match 7.6%; Score 79.5; DB 3; Length 3782;
Best Local Similarity 19.8%; Pred. No. 1.3e+02;
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;
QY 17 ALALAGFATQSPNSFYATGTAQAVSEPIDVESHLSITPAAGAGSGDDIGYAIWVI 75
Db 1552 AVRATGGYATVQSGSHTEAGQVYGVKSSPLLKLTGTHKLTISGFAMSNATP;LSLGLGW 1611
QY 76 KDQVNDVKLVTURNABQLKPYFYKIQITSGYETNSTAL-GNFSKAVISLDNPSAV 134
Db 1612 TPAADATIAKAVESARKARTAVVFAYDGTGVDRPNLSLPGTQDKLISAVADANPTI 1671
QY 135 IVLDKEDIAVLVP--DKTGYTNTSIWPG-----EPDKIIIVYNETKPVAILNFKAIFYEAK 198
Db 1672 VLVNTGS-SVLMFWLSKTRAV-LDMWYPGQAGAEATAALYGDVNP-----S 1716
QY 189 GMLFDSLPIVFN 200
Db 1717 GKLTQSFPAAEN 1728

RESULT 12
US-09-224-024-28
; Sequence 28, Application US/09224024
; Patent No. 6056953
; GENERAL INFORMATION:
; APPLICANT: Leslie Hickie
; APPLICANT: Jewel Payne
; TITLE OF INVENTION: Materials and Methods for the Control of
; TITLE OF INVENTION: Calliphoridae Pests
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/224,024
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/856,226
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA79
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-224-024-28

Query Match 7.6%; Score 79; DB 3; Length 1180;
Best Local Similarity 22.2%; Pred. No. 23;
Matches 47; Conservative 31; Mismatches 66; Indels 68; Gaps 10;
QY 24 FATTQS---PLNSFYATGTAQAVSEPIDVESHLSITPAAGAGSGDDIGYAIWIKDQVN 80
Db 520 FAWTHSSVDPKNTIYTHLTQI---PAVKANSIG--TASKVVQPGCHTGGDLIDFKDH-- 572
QY 81 DVKLAVTLRNABQLKPYFYKIQITSGYETNSTALGNFSETKAVISLONP----- 131
Db 573 ---FKITQHSNFOQSYF-----IRIRYASGSA-----NTRAVINLSIPGVVAELGMAL 618
QY 132 -----SAVIVLDKEDIAVLVPDKTGYTNTSIWPGCEPKIIV 168
Db 619 NTFSGSTDYTNLYKXDFQYLFESNEKVPAPNQNISLVFNRSVYNTTTLVI-----DKI-- 672
QY 169 YNETKPVAILNFKAIFYEAKGMLFDSLPIVFN 200
Db 673 --EFLPIT---RSIREDRKQKLETVOQIIN 698

RESULT 13
PCT-US94-07902-28
; Sequence 28, Application PC/TUS9407902
; GENERAL INFORMATION:
; APPLICANT: Street address: 4980 Carroll Canyon Road
; APPLICANT: City: San Diego
; APPLICANT: State/Province: California
; APPLICANT: Country: US
; APPLICANT: Postal code/zip: 92121
; APPLICANT: Phone number: (619) 453-8030
; APPLICANT: Telex number:
; TITLE OF INVENTION: Materials and Methods for the Control of
; TITLE OF INVENTION: Calliphoridae Pests
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07902
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA79
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-07902-28

Query Match 7.6%; Score 79; DB 5; Length 1180;
Best Local Similarity 22.2%; Pred. No. 23;
Matches 47; Conservative 31; Mismatches 66; Indels 68; Gaps 10;
PCT-US94-07902-28

```
QY 24 FATTQS---PLNSFYATGTAQVSEPIDVESHLSITPAAGAGSDDIGVAIVWIKDQVN 80
Db 520 FATHSSVDKNVIYTHLTQI---PAVKANSIG--TAKVVQPGHGTGGDLIDFKDH-- 572
QY 81 DVKLKVLTRNAEQKPYFKYLIQITSGYETNSFALGNFSETKAVISLDNP----- 131
Db 573 ---FKITCOHSNFQSYF-----IRIRYASNGSA-----NTRAVINLSIPGVAELGMAL 618
QY 132 -----SAVIVLDKEDIYLYPKDTGYTNTSIWVPGEPDKIIV 168
Db 619 NPFTSGDYTNLYKDKFYLEFSNVKFAPNQNISLVNRSVDVYNTVTILI---DKI-- 672
QY 169 YNETKPVAILNFKAFYEAKEGMLFDSLPLVFN 200
Db 673 --BELPIT---RSIREDKQKLETVQQIIN 698

RESULT 14
US-09-489-039A-12174
; Sequence 12174, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12174
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12174

Query Match 7.5%; Score 78.5; DB 4; Length 355;
Best Local Similarity 25.2%; Pred. No. 4;
Matches 57; Conservative 33; Mismatches 85; Indels 51; Gaps 13;

QY 6 LAIAGIIASAALALLAGFATTQSPINSFYA---TGTAQVSEPIDVE-SHLGSITPAAG 61
Db 36 LTVAALALSGSALAEVKIALVAKSLGNGFFFAANVGAQAQAKELGDKVVIYTGPTTTAE 95
QY 62 -----AQGSDDIGVAIVWIKDQVNDVKLV-TLRNAEQ-----LKPVEK 99
Db 96 AQIDVLNGLIAQGVDAIAIS-----ANDPDVVPVLKAMQGIKVVSWDSGVAPAGR 148
QY 100 YLQITSGYETNSFALGNFSETKAVISLDNPNSAVIVLDKEDIYLYPKDTGYTNTSIWV 159
Db 149 QIHLN-----PSNNALIG---ETNVKLAADALQALNV-EKGEVAVLSATPTS-INQNIWI 198
QY 160 PGEFDKIIVNETKPVAI-----INFKAFYEAKEGML--FDSLPLVI 198
Db 199 EEMKKVLPQVPSVQLVTVAYGDDLSDKSRYREA-VGLLKSYPDLKVI 243

RESULT 15
US-09-489-039A-13379
; Sequence 13379, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13379
```

```
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-13379

Query Match 7.5%; Score 78; DB 4; Length 408;
Best Local Similarity 22.2%; Pred. No. 5.7;
Matches 69; Conservative 31; Mismatches 73; Indels 138; Gaps 17;

QY 1 VKYT---TIAIAGIIASAALALLAGFATTQSPINSFYATGTAQVSEPIDVESHU--- 53
Db 49 MKYTLPALTLAIS-----AALSGCATPHS-----SAVSQPV-VDSPVPNV 87
QY 54 -----GSITPAAGAGSDDIGVAIVWIKD--QVNDVKLVKT-LRN 90
Db 88 AQPLQRLAELYEMALSPQGDALYVASAEGFKVQGGAVITLDPHTLTIGLTHDOLKN 147
QY 91 -AEQLKPYFK--YLQITSGYETNSTALGN-----FSE-----TKAVISLDN-- 130
Db 148 FALQLSAEGKTLVSNSLDGGISAIDTATGKVKNRLLFSENEKGRPYGARQLLLNNTL 207
QY 131 -----PSAVIVLDKEDIYLYPKDTGYTNTSIWVPG-----EPDKIIVY 169
Db 208 YVGAVADPPAQIWIYVDATTLKL-----KTRIKNTGKWMTGLHYSAGTGRVVAANGSEILVI 263
QY 170 NE-----TKPVAILNFKAFYEAKEGMLF----- 192
Db 264 NPNQRIEQRWKPLGDKPALLNMAE--DSDTGRLFVTDNSKAKTTLVLDIHSGKLLKQL 321
QY 193 ---DSLPLVFN 200
Db 322 DVGDSLAVQFN 332

Search completed: April 6, 2004, 12:33:56
Job time : 23 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:33:31 ; Search time 42 Seconds
(without alignments)
1294.407 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTLAIAGIIASAAALAL.....EGMLEDSPVIFNFVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1042	100.0	207	10	US-09-997-807-2
2	1042	100.0	207	15	US-10-370-370-2
3	553.5	53.1	170	10	US-09-997-807-4
4	553.5	53.1	170	15	US-10-370-370-4
5	540	51.8	178	10	US-09-997-807-6
6	540	51.8	178	15	US-10-370-370-6
7	473	45.4	140	15	US-10-370-370-12
8	307.5	29.5	131	15	US-10-370-370-8
9	305.5	29.3	124	10	US-09-997-807-10
10	305.5	29.3	124	15	US-10-370-370-10
11	304.5	29.2	130	10	US-09-997-807-8
12	96	9.2	255	14	US-10-134-021-4
13	82	8.8	322	12	US-10-282-122A-61228
14	86.5	8.3	846	10	US-09-893-519A-34
15	86.5	8.3	846	15	US-10-369-493-1972

16	85.5	8.2	337	12	US-10-412-699B-1672
17	85.5	8.2	337	15	US-10-374-780A-1646
18	85.5	8.2	845	12	US-10-072-012-415
19	84.5	8.1	506	15	US-10-108-260A-4387
20	84.5	8.1	1886	12	US-10-147-299A-6
21	83	8.0	626	9	US-09-948-722-2
22	83	8.0	1246	9	US-09-919-497-85
23	83	8.0	1247	10	US-09-961-403-14
24	82.5	7.9	243	14	US-10-171-404A-42
25	82.5	7.9	243	15	US-10-374-780A-2004
26	82.5	7.9	296	12	US-10-282-122A-52171
27	82.5	7.9	1954	12	US-10-147-299A-4
28	81.5	7.8	320	12	US-10-335-977-5813
29	81.5	7.8	349	12	US-10-335-977-5814
30	81	7.8	530	12	US-10-424-599-262050
31	81	7.8	751	14	US-10-032-585-7803
32	80.5	7.7	258	9	US-09-738-626-6436
33	80.5	7.7	292	12	US-10-424-599-240558
34	80	7.7	398	12	US-10-282-122A-48705
35	80	7.7	612	12	US-10-282-122A-64164
36	80	7.7	1647	14	US-10-032-585-7572
37	79.5	7.6	317	15	US-10-369-493-6984
38	79.5	7.6	447	12	US-10-282-122A-51512
39	79.5	7.6	769	10	US-09-793-708-12
40	79.5	7.6	769	14	US-10-201-365-10
41	79.5	7.6	769	14	US-10-160-539-12
42	79.5	7.6	809	9	US-09-861-289-24
43	79.5	7.6	809	9	US-09-860-846-24
44	79.5	7.6	809	10	US-09-988-384B-24
45	79.5	7.6	809	10	US-09-836-821-24

ALIGNMENTS

RESULT 1

US-09-997-807-2
; Sequence 2, Application US/09997807
; Publication No.: US20030198661A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; TITLE OF INVENTION: Uses of the Polymer
; FILE REFERENCE: DVSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
; US-09-997-807-2

Query Match 100.0%; Score 1042; DB 10; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.8e-101;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTLAIAGIIASAAALALAGFATTQSPPLNSFYATGTAQVSEPIDVESHGSIPTAA 60

Db 1 VKYTLAIAGIIASAAALALAGFATTQSPPLNSFYATGTAQVSEPIDVESHGSIPTAA 60

QY 61 GAGGSDDIGYAIWIKQDQNDVKLVKTLRNASQLPKYFYKIQITSGYETNSTALGNFS 120

Db 61 GAGGSDDIGYAIWIKQDQNDVKLVKTLRNASQLPKYFYKIQITSGYETNSTALGNFS 120

QY 121 ETXAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVPGPKIIVYNETKPVAILNF 180

Sequence 1672, Ap
Sequence 1646, Ap
Sequence 415, App
Sequence 4387, Ap
Sequence 6, Appli
Sequence 2, Appli
Sequence 85, Appl
Sequence 14, Appl
Sequence 42, Appl
Sequence 2004, Ap
Sequence 52171, A
Sequence 4, Appli
Sequence 5813, Ap
Sequence 5814, Ap
Sequence 262050,
Sequence 7803, Ap
Sequence 6436, Ap
Sequence 240558,
Sequence 48705, A
Sequence 64164, A
Sequence 7572, Ap
Sequence 6984, Ap
Sequence 51512, A
Sequence 12, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 24, Appl

Db 121 ETKAVISLNDPSAVIVLDKEDIAVLDPDKGTNTSIWVGPBPKIIVVNETKPVALLNF 180
QY 181 KAFYEAKGMLFDSLPLVIFNFQVLQVG 207
Db 181 KAFYEAKGMLFDSLPLVIFNFQVLQVG 207

RESULT 2
US-10-370-370-2
; Sequence 2, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-10-370-370-2

Query Match 100.0%; Score 1042; DB 15; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.8e-101;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VKYTTLAIAGIIASAALALLAGFATTQSPPLNSFYATGTAAQAVSEPIDVESHLSITPAA 60
Db 1 VKYTTLAIAGIIASAALALLAGFATTQSPPLNSFYATGTAAQAVSEPIDVESHLSITPAA 60
QY 61 GAQGSDDIGYAIWIKQDVNDVKLKTNRNAEOLKPYFKYLOIQTSGYETNSTALGNFS 120
Db 61 GAQGSDDIGYAIWIKQDVNDVKLKTNRNAEOLKPYFKYLOIQTSGYETNSTALGNFS 120
QY 121 ETKAVISLNDPSAVIVLDKEDIAVLDPDKGTNTSIWVGPBPKIIVVNETKPVALLNF 180
Db 121 ETKAVISLNDPSAVIVLDKEDIAVLDPDKGTNTSIWVGPBPKIIVVNETKPVALLNF 180
QY 181 KAFYEAKGMLFDSLPLVIFNFQVLQVG 207
Db 181 KAFYEAKGMLFDSLPLVIFNFQVLQVG 207

RESULT 3
US-09-997-807-4
; Sequence 4, Application US/09997807
; Publication No. US20030198681A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE REFERENCE: DUSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 170

; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-4

Query Match 53.1%; Score 553.5; DB 10; Length 170;
Best Local Similarity 60.2%; Pred. No. 4.4e-50;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;

QY 1 VKYTTLAIAGIIASAALALLAGFATTQSPPLNSFYATGTAAQAVSEPIDVESHLSITPAA 60
Db 1 VKPTALALAGIIASAADLALLAGFATTQSPPLNSFYATGTAAATSEPIDVESHLSITPAA 60
QY 61 GAQGSDDIGYAIWIKQDVNDVKLKTNRNAEOLKPYFKYLOIQTSGYETNSTALGNFS 120
Db 61 GAQGSDDIGYAIWIKQDVNDVKLKTNRNAEOLKPYFKYLOIQTSGYETNSTALGNFS 120
QY 121 ETKAVISLNDPSAVIVLDKEDIAVLDPDKGTNTSIWVGPBPKIIVVNETKPVALLNF 180
Db 112 ETKAVISIDKPSAVIILDSQDF-----DSNNRAKISA 143
QY 181 KAFYEAKGMLFDSLPLVIFNFQVLQV 206
Db 144 TAYEAKGMLFDSLPLIFNIQVLSV 169

RESULT 4
US-10-370-370-4
; Sequence 4, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Frey, Gerhard
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-10-370-370-4

Query Match 53.1%; Score 553.5; DB 15; Length 170;
Best Local Similarity 60.2%; Pred. No. 4.4e-50;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;

QY 1 VKYTTLAIAGIIASAALALLAGFATTQSPPLNSFYATGTAAQAVSEPIDVESHLSITPAA 60
Db 1 VKPTALALAGIIASAADLALLAGFATTQSPPLNSFYATGTAAATSEPIDVESHLSITPAA 60
QY 61 GAQGSDDIGYAIWIKQDVNDVKLKTNRNAEOLKPYFKYLOIQTSGYETNSTALGNFS 120
Db 61 GAQGSDDIGYAIWIKQDVNDVKLKTNRNAEOLKPYFKYLOIQTSGYETNSTALGNFS 120
QY 121 ETKAVISLNDPSAVIVLDKEDIAVLDPDKGTNTSIWVGPBPKIIVVNETKPVALLNF 180
Db 112 ETKAVISIDKPSAVIILDSQDF-----DSNNRAKISA 143
QY 181 KAFYEAKGMLFDSLPLVIFNFQVLQV 206
Db 144 TAYEAKGMLFDSLPLIFNIQVLSV 169

RESULT 5
US-09-997-807-6
; Sequence 6, Application US/09997807
; Publication No. US20030198681A1

```
/ GENERAL INFORMATION:
/ APPLICANT: Jay Short
/ APPLICANT: Eric J. Mathur
/ APPLICANT: W. Michael Lafferty
/ APPLICANT: Nelson Barton
/ APPLICANT: Kevin Chow
/ TITLE OF INVENTION: Method of Making A Protein Polymer and
/ FILE REFERENCE: DUSA-1005US
/ CURRENT APPLICATION NUMBER: US/09/997,807
/ CURRENT FILING DATE: 2001-11-30
/ PRIOR APPLICATION NUMBER: 60/250,426
/ PRIOR FILING DATE: 2000-11-30
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 6
/ LENGTH: 178
/ TYPE: PRT
/ ORGANISM: Pyrodictium abyssi
US-09-997-807-6

Query Match
Best Local Similarity 51.8%; Score 540; DB 10; Length 178;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY 1 VKYTTLAIAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSGITPA 59
Db 1 MRYTTLALAGIVASAAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLDNTIAPA 60

QY 60 AGAQSDDIGYAIWIKDQVNDVKLTARNABOLKPYFKYLIQITSGYETNSTALGNF 119
Db 61 AGAQGYKDMGYIKITQSKVNVIKLTARNABOLKPYFDYQLVLTS-----NATGT--- 113

QY 120 SETKAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILN 179
Db 114 DMVKAVLSLEKPSAVILNDND-----YDSTNKIQ-LK 145

QY 180 FKAFYEAKGMLFDSLPIVFNFOVL 204
Db 146 VEAYEAKGMLFDSLPIVFNFOVL 170

RESULT 6
US-10-370-370-6
/ Sequence 6, Application US/10370370
/ Publication No. US20040006778A1
/ GENERAL INFORMATION:
/ APPLICANT: Barton, Nelson R.
/ APPLICANT: O'Donoghue, Eileen
/ APPLICANT: Frey, Gerhard
/ TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
/ FILE REFERENCE: 09010-115001
/ CURRENT APPLICATION NUMBER: US/10/370,370
/ CURRENT FILING DATE: 2003-02-18
/ PRIOR APPLICATION NUMBER: 60/357,406
/ PRIOR FILING DATE: 2002-02-15
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 6
/ LENGTH: 178
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ OTHER INFORMATION: consensus sequence
US-10-370-370-6

Query Match
Best Local Similarity 45.4%; Score 473; DB 15; Length 140;
Best Local Similarity 61.3%; Pred. No. 1e-41;
Matches 119; Conservative 7; Mismatches 8; Indels 60; Gaps 8;

QY 5 TLAIAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSGITPAAGAOG 64
Db 3 TLAIAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSGITPAAGAOG 59

QY 65 SDDIGYAIWIKDQVNDVKLTARNABOLKPYFKYLIQITSGYETNSTALGNFSETKA 124
Db 60 S-DIGYII-----KVNIVKLTARNABOLKPYFKYLIQIVLS-----SEIKA 100

QY 125 VISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILNFKAFY 184
Db 101 VISLDKPSAVIILD-EDFAI-----AYY 122

QY 185 EAKGMLFDSLPIV 198
Db 123 EAKGMLFDSLPIV 136

RESULT 8
US-10-370-370-8
/ Sequence 8, Application US/10370370
/ Publication No. US20040006778A1
/ GENERAL INFORMATION:
/ APPLICANT: Barton, Nelson R.
/ APPLICANT: O'Donoghue, Eileen
/ APPLICANT: Frey, Gerhard
/ TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
/ FILE REFERENCE: 09010-115001
/ CURRENT APPLICATION NUMBER: US/10/370,370
/ CURRENT FILING DATE: 2003-02-18
/ PRIOR APPLICATION NUMBER: 60/357,406
/ OTHER INFORMATION: consensus sequence
US-10-370-370-8

Query Match
Best Local Similarity 51.8%; Score 540; DB 15; Length 178;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY 1 VKYTTLAIAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSGITPA 59
Db 1 MRYTTLALAGIVASAAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLDNTIAPA 60

QY 60 AGAQSDDIGYAIWIKDQVNDVKLTARNABOLKPYFKYLIQITSGYETNSTALGNF 119
Db 61 AGAQGYKDMGYIKITQSKVNVIKLTARNABOLKPYFDYQLVLTS-----NATGT--- 113

QY 120 SETKAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILN 179
Db 114 DMVKAVLSLEKPSAVILNDND-----YDSTNKIQ-LK 145

QY 180 FKAFYEAKGMLFDSLPIVFNFOVL 204
Db 146 VEAYEAKGMLFDSLPIVFNFOVL 170

RESULT 6
US-10-370-370-6
/ Sequence 6, Application US/10370370
/ Publication No. US20040006778A1
/ GENERAL INFORMATION:
/ APPLICANT: Barton, Nelson R.
/ APPLICANT: O'Donoghue, Eileen
/ APPLICANT: Frey, Gerhard
/ TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
/ FILE REFERENCE: 09010-115001
/ CURRENT APPLICATION NUMBER: US/10/370,370
/ CURRENT FILING DATE: 2003-02-18
/ PRIOR APPLICATION NUMBER: 60/357,406
/ PRIOR FILING DATE: 2002-02-15
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 6
/ LENGTH: 178
/ TYPE: PRT
/ ORGANISM: Pyrodictium abyssi
US-10-370-370-6

Query Match
Best Local Similarity 51.8%; Score 540; DB 15; Length 178;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY 1 VKYTTLAIAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSGITPA 59
Db 1 MRYTTLALAGIVASAAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLDNTIAPA 60
```

```
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-10-370-370-8

Query Match      29.5%; Score 307.5; DB 15; Length 131;
Best Local Similarity 48.1%; Pred. No. 2.6e-24;
Matches 76; Conservative 16; Mismatches 39; Indels 27; Gaps 4;

Qy 33 SFYATGTAQAVSEPIDVESHLSITPAAGAGSGDDIGYAIWIKQDVNDVKLVTLRNAE 92
Db 1 SFYATGTAQAVSEPIDVSSLSLTNTAAGAGGKQTLGDIITYAHNDVNIITKLKVTLANAA 60

Qy 93 QLRPFYKYLQIQTSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIYVLPDKTG 152
Db 61 QLRPFYKYLILKLVSLDSN---GNESEKGMITLWKFYAVIILHEDF----- 105

Qy 153 TNSIWPGEPEDKIIIVNETKPVAILNFKAFYEAKGML 190
Db 106 -NNDIDNGND-----AKIRVYAYEAKGML 131

RESULT 9
US-09-997-807-10
; Sequence 10, Application US/09997807
; Publication No. US20030198681A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE REFERENCE: DVSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-10

Query Match      29.3%; Score 305.5; DB 10; Length 124;
Best Local Similarity 48.8%; Pred. No. 3.9e-24;
Matches 78; Conservative 14; Mismatches 31; Indels 37; Gaps 5;

Qy 33 SFYATGTAQAVSEPIDVESHLSITPAAGAGSGDDIGYAIWIKQDVNDVKLVTLRNA 91
Db 1 SFYATGTAQAVSEPIDVSSLSLTNTAIAAPAAGAGSGVIGSITIENKTDVNVVVKLITLANA 60

Qy 92 EQLKPYFKYLIQIQTSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIYVLPDKTG 151
Db 61 EQLKPYFDYLIQVLKSLDSN-----EIKAVLSLEKPSAVIILNEDF-----QG 104

Qy 152 YTNISWVPEPEDKIIIVNETKPVAILNFKAFYEAKGML 191
Db 105 GDNQ-----CQIDATAYEAKGML 124

RESULT 11
US-09-997-807-8
; Sequence 8, Application US/09997807
; Publication No. US20030198681A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE REFERENCE: DVSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-8

Query Match      29.2%; Score 304.5; DB 10; Length 130;
Best Local Similarity 47.2%; Pred. No. 5.4e-24;
Matches 75; Conservative 22; Mismatches 33; Indels 29; Gaps 6;

Qy 33 SFYATGTAQAVSEPIDVESHLSITPAAGAGSGDDIGYAIWIKQDVNDVKLVTLRNAE 92
Db 1 SFYATGTAQAVSEPIDVSSLSLTNTAAGAGGKQTLGDIITYAHNDVNIITKLKVTLANAA 60

Qy 93 QLRPFYKYLQIQTSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIYVLPDKTG 152
Db 61 QLRPFYKYLILKLVSLDSN---GNESEKGMITLWKFYAVIILHEDF-----F 105

Qy 153 TNSIWPGEPEDKIIIVNETKPVAILNFKAFYEAKGML 191
```


Db 213 IPNFESGLHLIFEA--GLLYQSLGYKVEKFRMLNI 246

Search completed: April 6, 2004, 12:39:46
Job time : 42 secs

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 107
US-09-893-519A-34

Query Match      8.3%; Score 86.5; DB 10; Length 846;
Best Local Similarity 24.1%; Pred. No. 8.5;
Matches 52; Conservative 38; Mismatches 81; Indels 45; Gaps 12;

QY 25 ATTQSPLNSFYATGTAQAVSEPI---DVESHLSGITPAAGAGSDD-IGYAIWIKQOVN 80
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 ATQLGPLIDFYKLQSLDSPETIMWHKIEKFDAL---FGIQTDDMVKLSVFQSLPS 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 DVKLKVTLR---NAEQKPYFKYLQ-IQITSGY-ETNSTALGNFSETKAVIS----- 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 NYRAKIVQKSSGLNMLNLANHEHLLSPVRAPSIYTEASFENMDRFSRRSMVSPNRYVP 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 -----LDNPSAVILDKEDIAVLDPKTYNTSIWVGPEDPKIIVYNETKPA 176
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 SSTYSSVTLRQLSNPYVYNTIPEEDI-LKYVSYTLTATTSALPPFDHEQIQI-----PSK 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 ILNFKA-----FYEAKGMLFDSLPV-IFNFQVLQV 206
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 IPNFESGLHLIFEA--GLLYQSLGYKVEKFRMLNI 246
```

```
RESULT 15
US-10-369-493-1972
; Sequence 1972, Application US/10369493
; Publication No. US2003033675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1972
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1972
```

```
Query Match      8.3%; Score 86.5; DB 15; Length 846;
Best Local Similarity 24.1%; Pred. No. 8.5;
Matches 52; Conservative 38; Mismatches 81; Indels 45; Gaps 12;

QY 25 ATTQSPLNSFYATGTAQAVSEPI---DVESHLSGITPAAGAGSDD-IGYAIWIKQOVN 80
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 ATQLGPLIDFYKLQSLDSPETIMWHKIEKFDAL---FGIQTDDMVKLSVFQSLPS 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 DVKLKVTLR---NAEQKPYFKYLQ-IQITSGY-ETNSTALGNFSETKAVIS----- 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 NYRAKIVQKSSGLNMLNLANHEHLLSPVRAPSIYTEASFENMDRFSRRSMVSPNRYVP 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 -----LDNPSAVILDKEDIAVLDPKTYNTSIWVGPEDPKIIVYNETKPA 176
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 SSTYSSVTLRQLSNPYVYNTIPEEDI-LKYVSYTLTATTSALPPFDHEQIQI-----PSK 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 ILNFKA-----FYEAKGMLFDSLPV-IFNFQVLQV 206
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:37:42 ; Search time 20 Seconds
(without alignments)
995.582 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTTLAAGIATASAAAL.....EGMLFDSLPLVFNQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.5	9.2	262	T22599	1-acylglycerol-3-p
2	91	8.7	427	G69067	N-ethylamine ch
3	91	8.7	533	G72593	hypothetical prote
4	90	8.6	174	B83337	hypothetical prote
5	89.5	8.6	449	JC7306	extracellular prot
6	88.5	8.5	334	A71169	probable flagellin
7	88.5	8.5	1148	S72635	exo-poly-alpha-gal
8	88	8.4	378	A90280	hypothetical prote
9	86.5	8.3	284	JC5221	stomatin - mouse
10	86.5	8.3	443	F90167	conserved hypothet
11	86.5	8.3	846	S59262	hypothetical prote
12	86	8.3	1258	JQ0188	ice nucleation pro
13	85.5	8.2	329	IMBXEB	immunogenic protei
14	85.5	8.2	329	AF3351	3ik immunogenic pr
15	84.5	8.1	223	S74812	ABC-type transport
16	84.5	8.1	739	B82552	phage-related prot
17	84	8.1	391	B70625	probable PPE prote
18	83.5	8.0	468	1 VGBBHH	glycoprotein gp13
19	83.5	8.0	468	1 B46114	glycoprotein gp13
20	83.5	8.0	870	G69006	DNA-directed RNA p
21	83.5	8.0	4436	E71086	hypothetical prote
22	83	8.0	1247	1 MMHUND	nidogen precursor
23	83	8.0	1322	S07053	ice nucleation pro
24	82.5	7.9	243	S58494	auxin-induced prot
25	82.5	7.9	236	D97052	homoserine kinase
26	82.5	7.9	402	G90258	conserved hypothet
27	82	7.9	147	2 D69008	conserved hypothet
28	82	7.9	350	2 A48421	ornithine transcar
29	82	7.9	1942	2 B71426	hypothetical prote

septum formation p
laccase (EC 1.10.3
RGPI protein - yea
probable methylmal
fibrinogen-binding
hypothetical prote
hypothetical prote
probable peptidogl
hypothetical prote
HlyD family secret
hypothetical prote
hypothetical prote
glucose inhibited
hypothetical prote
fima protein - Dic
C4-dicarboxylate t
site-specific DNA-

ALIGNMENTS

RESULT 1

T22599

1-acylglycerol-3-phosphate O-acyltransferase (EC 2.3.1.51) homolog F59P4.4 - Caenorhabd
C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 06-Oct-2000

C:Accession: T22599; T23012

R:Dobson, R.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19587

A:Accession: T22599

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-262 <WIL>

A:Cross-references: EMBL:Z81089; PIDN:CA803139.1; GSPDB:GN00028; CESP:F59P4.4

A:Experimental source: clone F53H4

R:Harris, B.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19652

A:Accession: T23012

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-262 <W12>

A:Cross-references: EMBL:Z81095; PIDN:CA803160.1; GSPDB:GN00028; CESP:F59P4.4

A:Experimental source: Clone F59F4

C:Genetics:

A:Gene: CESP:F59P4.4

A:Map position: X

A:Introns: 57/2; 102/1; 160/3; 192/3

C:Superfamily: mouse 1-acylglycerol-3-phosphate O-acyltransferase

C:Keywords: acyltransferase; coenzyme A

Query Match 9.2%; Score 95.5; DB 2; Length 262;

Best Local Similarity 24.2%; Pred. No. 1.9;

Matches 58; Conservative 27; Mismatches 96; Indels 59; Gaps 9;

QY 4 TTUAIAGIATASAAALA-----LLAGPATT---QSPLNSFYAG 38
DB 2 TFLAILFVIAVLLLAQLPVGIFYRAVYFGMCLITGGFLGLGIASIPFGKSPNNHPRMFK 61
QY 39 TAQAVSEPIDVESHLGS-----ITPAAGAQSGDDIGVAIVIKQVNDVKLVTL 88
DB 62 IFQMTWPGVGFELNSILHDKRPYIIIANHQSLDVLGMSFAMPVDCV--VMLKSSL 119
QY 89 RNAEQLPKYFKLQIQITSGYETNSTALGNFSETKRAVILSDNPSAVIVLDKEDIAVLYPD 148
DB 120 -----KYLPGNLCAYLDCSVYINRFSEKALKATVDTTLHIVTKRKVWI-YPE 168
QY 149 KTGTYNTISWPGEPDKIIVYNETK-----PVALINFKAFYEAKB-----GMLFDSLPLVI 198
DB 169 GTRNABPEL-LPFPKGAFLAKQAKIPVPCVFSHKFFYSHAERKLTSGNCIIDLPEV 227

Db 68 LISAAGIPATLQGSIEASLNGTSFADVADAKDVTLLTDGKGVAVSYDKNTLTKLSD 127
 QY 103 IQITSGYETNSTALGNFSSTKA-----VISLDNPSAVIV-----LDKEDIAVLVPDKTYGYN 154
 Db 128 VRAGDDYTWTLSGVG-FSGKANAGKTLIFKLPBGVTVVEGANYNKDDHKVTLL-DQYGNVS 185
 QY 155 TSIVWPGEPDKIIVNETKPVAILNPKAFYEAKGML 191
 Db 186 GLKFFV--ISKVKAYDSANTNAV-----SFYDAKSGLV 215
 RESULT 6
 A71169
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
 R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi
 DNA Res. 5, 55-76, 1998
 A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
 A;Reference number: A71000; PMID:98344137; PMID:9679194
 A;Accession: A71169
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-334 <RAW>
 A;Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29638.1; PID:g3256955
 A;Experimental source: strain OT3
 A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C;Genetics:
 A;Gene: PH0549
 C;Superfamily: archaeal flagellin
 Query Match 8.5%; Score 88.5; DB 2; Length 334;
 Best Local Similarity 22.1%; Pred. No. 9.6;
 Matches 42; Conservative 30; Mismatches 63; Indels 55; Gaps 8;
 QY 8 IAGIIASAAALALL-----AGFATTQSPNLSFYAT---GTAQAVSEPIDVES 51
 Db 15 IAWLVAAVAVALINTSGYLQKQSQTGRQTTQEVASGIKVTRVWGKDSATNPTVIOE 74
 QY 52 HLGSIITPAAGQSGDDIGYALVWTKQVNDVKLVKVTLRNAEQLKPYKYLQITSGYET 111
 Db 75 LAVYITENAGSSG-----IDLTKVRITLSDG-QKQAIKPY----- 108
 QY 112 NSTALGNFSSTKAIVS--LDNPSAVIVLDKEDIAVLVPDKTYGYNNTSIVWPGEPDKIIVY 169
 Db 109 ---RVGNSANELYFLAELMQLGATIKFDNGSVQV-YFDPDWTISAA-----PTVID 157
 QY 170 NETKPVAILN 179
 Db 158 TTNKVBIVN 167
 RESULT 7
 S72635
 C:Species: Thermoanaerobacterium thermosulfurigenes
 C:Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 15-Oct-1999
 C;Accession: S72635; S72621
 R;Matuschek, M.; Sahm, K.; Bahl, H.
 submitted to the EMBL data library, March 1996
 A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes B
 A;Reference number: S72635
 A;Accession: S72635
 A;Molecule type: DNA
 A;Residues: 1-1148 <MATU>
 A;Cross-references: EMBL:U0951; NID:g1542972; PIDN:AAB08040.1; PID:g1255234
 A;Experimental source: strain EM1
 R;Matuschek, M.; Sahm, K.; Zibat, A.; Bahl, H.
 Mol. Gen. Genet. 252, 493-496, 1996
 A;Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 th

A;Reference number: S72621; PMID:97033555; PMID:8879252
 A;Accession: S72621
 A;Molecule type: DNA
 A;Residues: 729-1148 <MAW>
 A;Cross-references: EMBL:U0951
 A;Experimental source: EM1
 C;Genetics:
 A;Gene: P91A
 C;Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
 C:Keywords: A cellulose-binding repeat homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F;1-30/Domain: signal sequence #status predicted <SIG>
 F;31-1148/Product: exo-poly-alpha-galacturonosidase #status predicted <MAT>
 F;969-1022/Domain: S-layer repeat homology <SLR1>
 F;1028-1081/Domain: S-layer repeat homology <SLR2>
 F;1092-1144/Domain: S-layer repeat homology <SLR3>
 Query Match 8.5%; Score 88.5; DB 2; Length 1148;
 Best Local Similarity 19.1%; Pred. No. 46;
 Matches 37; Conservative 38; Mismatches 82; Indels 37; Gaps 6;
 QY 20 LLAGFATTQSPNLSFYATGTAQAVSEPIDVESHLGSIITPAAGQSGDDIGYALVWIK-DQ 78
 Db 14 LLAFLMVFTMPMSKAFADTTSSGPAAPVLDQLVLSGVT-----DTGFTLVVHKPDN 64
 QY 79 VNDV-KLVTLRNAEQLKPYKYLQITSGY-----ETNSTALGNF--SE 121
 Db 65 YSDITDYKTIVSDSVYEQYIYASENQTVASQYIKQYDNNVGDLDKDDNGTVNSAYKISM 124
 QY 122 TKAVISLDNPSAVIVLDKEDIAVLVPDKTYGYNNTSIVWPGEPDKIIVNETKPVAILNFK 181
 Db 125 HSFVLGLKPTLYTTQVSDANKNTSTPTVITQSTAPTPSE-----NIINVE 174
 QY 182 AFYEAKEGMLFDLSL 195
 Db 175 STGAVGDGVLADDV 188
 RESULT 8
 A90280
 C:Species: Sulfolobus solfataricus
 C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C;Accession: A90280
 R;She, Q.; Singh, R.K.; Confalonieri, F.; Zhivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
 Jong, I.; Jeffries, A.C.; Kozera, C.W.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A;Description: Sulfolobus solfataricus complete genome.
 A;Reference number: A99139
 A;Accession: A90280
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-378 <KUR>
 A;Cross-references: GB:AE006641; NID:g13814452; PIDN:AAK41496.1; GSPDB:GN00155
 C;Genetics:
 A;Gene: SSO1258
 Query Match 8.4%; Score 88; DB 2; Length 378;
 Best Local Similarity 22.5%; Pred. No. 12;
 Matches 47; Conservative 31; Mismatches 67; Indels 64; Gaps 10;
 QY 19 ALLAGFATTQSPNLSFYATGTAQAV-----SEPIDVESHLGSIITPAAG----AOGSD 66
 Db 43 ALVAG-----SGIINSYLSIINDIAPLLSRVLSSEPIDVITILEKILFSAGNCVWSGAI 98
 QY 67 DIGYALVW-----IKDVND-VKL-----KVTLRNAEQLKPYFKY 100
 Db 99 SAVEMALWLSKARKSNVELKLGDKTRDSVKVVASPRFGKIDDDVLIATKSLERGFDL 158
 QY 101 LQIQITSGYETNSTALG-----NFSSETKAVISLDNPSAVIVLDKEDIAVLVPDKT--- 150
 Db 159 VKL-----HQSPSTVLAAVKAIKENYKVKIADLNSP-----FDNLDVAKFEVDKVKHY 208

[illegible]

RESULT 12

JQ0188
ice nucleation protein - Erwinia herbicola
C/Species: Erwinia herbicola
C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 26-Aug-1999
C/Accession: JQ0188; S35273
R/Warren, G.; Corotico, L.
Gene 85, 239-242, 1989
A/Title: The consensus sequence of ice nucleation proteins from Erwinia herbicola, Pseudomonas fluorescens Pf-0, and Pseudomonas fluorescens Pf-5
A/Reference number: JQ0188; MUID: 90152370; PMID: 2515997
A/Accession: JQ0188
A/Molecule type: DNA
A/Residues: 1-1258 <WAR>
A/Cross-references: EMBL:M26382; NID:G148419; PIDN:AAA24823.1; PID:G148420
R/Experimental source: strain M1
R/Gurian-Sherman, D.; Lindow, S.E.; Panopoulos, N.J.
Mol. Microbiol. 9, 383-391, 1993
A/Title: Isolation and characterization of hydroxylamine-induced mutations in the Erwinia
A/Reference number: S35273; MUID: 94018633; PMID: 8412688
A/Accession: S35273
A/Status: preliminary
A/Molecule type: protein
A/Residues: 387-536 <GUR>
C/Genetics:
A/Gene: iceE
C/Superfamily: ice nucleation protein

Query Match	8.3%	Score 86;	DB 2;	Length 1258;
Best Local Similarity	23.7%;	Pred. No. 82;		
Matches	47;	Conservative 26;	Mismatches 65;	Indels 60; Gaps 11.
Qy	3	YTTTAAIGITASAAALALLAGFAATTQSP-LNSFVAT--GTAQA VSEPIDVESHLSGIITPA	59	
Db	676	YGSNSTAG-----ADSSLIAGYGSTQTAGYNISLLTAGYGSTQTAGEGSDLTSGYGS-TST	729	
Qy	60	AGAAGSDDIGVAIWIKDQNDVKIKVTLRNAEQLKPYFKYL-----QIQITS	107	
Db	730	AGADSSLIAGY-----GSTQTASYHSLLTAGYGYSTQTAREQSIVLTT	770	
Qy	108	GYTEYNSTA-----LGNFSETK----AVISLDNPSAVIDUKEDIAVLYPDKTGYNTS	156	
Db	771	GYGSTSIAGDSSLIAGYGYSTQTAGYHSLIITAGYGYSTQTAEGRSDLT-----TGCGSTS	824	
Qy	157	IWVPGEPPKIIV-YNETK	173	
Db	825	--TAGADSSLIAGYGSTQT	840	

RESULT 13
IMBKBB
immunogenic protein BCSP31 precursor - Brucella abortus
C/Species: Brucella abortus
C/Date: 30-Jun-1989 #sequence revision 30-Jun-1989 #text change 16-Jul-1999

C;Accession: J02079
R;Wayfield, J.E.; Bricker, B.J.; Godfrey, H.; Crosby, R.M.; Knight, D.J.; Halling, S.M.
Gene 63, 1-9, 1988
A;Title: The cloning, expression, and nucleotide sequence of a gene coding for an immun
A;Reference number: J02079; MUID:88255848; PMID:3133283
A;Accession: J02079
A;Molecule type: mRNA
A;Residues: 1-329 <MAT>
A;Cross-references: GB:M20404; NID:G144104; PIDN:AAA22993.1; PID:G144105
C;Comment: Brucella abortus is the causative agent for brucellosis in cattle and man.
C;Superfamily: Immunogenic protein BCSP31
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-329/Product: immunogenic protein BCSP31 #status predicted <MAT>

```

Query Match      8.2%; Score 85.5; DB 1; Length 329;
Best Local Similarity 19.3%; Pred. No. 17;
Matches          56; Conservative 28; Mismatches 73; Indels 133; Gaps 13;

QY   2 KYTTLAAGIIASAAALALLAGFATTQTSPNSFYATGTAQAVSEPIDVESHGLSGITPAAG 61
    |||:|||:|||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   6 KIRRLAAVAV---AGAATAGASFAVAQAP--TFTRIGTGCTACTYYPIGGLIANAISGAG 60

QY   62 AQG-----SDDIGYAIVMI-----KDOVND 81
    :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   61 EKGUGLIVATAVSSNGSVANINAIKGALESGFTQSDVAY---WAYNGTGLYDGKGKVED 117

QY   82 VKLKVTILR-----SDDIGYAIVMI-----KDOVND 89
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   118 LRLLATIYPETIHIVARKDANIKSVDLKGKRVSLDEPGSGTIVDARIVLEAYGLTEDDI 177

QY   90 NAOQLKP-----YFKYLQIQITSGVETNSTALGNPSETKAVISLDNPSAVI 135
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   178 KAEHLFGPAGERIKDQALDAYF-----FWGGYPTGAISELAINSGISILVPIISGPPEADK 231

QY   136 VLDK-----EDI--AVLYPKTYNTSI---WVPG--EPDKLIIVNETK 173
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   232 ILEKYSFFSKDVVPAGAKOVAETPTLLUAAQAQWVTSAKQPDDLLI-YNIHK 280

```

RESULT 14
AF3351
31K immunogenic protein precursor [imported] - *Brucella melitensis* (strain 16M)
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AF3351
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.;
Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagijs, S.; O'Callaghan, D.;
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella*
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3351
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51977.1; PID:g17982738; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0796
A:Map position: 1
C:Superfamily: immunogenic protein BCSP31

Query Match	8.2:	Score 85.5;	DB 2:	Length 329;
Best Local Similarity	19.3%;	Pred. No. 17;		
Matches	56;	Conservative	28;	Mismatches
			73;	Indels 133;
				Gaps 13;

```
QY      82 VKLKVTLR----- 89
DB      118 LRLALYLPETIHIVARKDANIKSVADLKGKRVSLDEPGSGITIVDARIIVLEAYGLTEDDI 177
QY      90 NABOLKP-----YFKYLIQITSGYETNSTALGNFSETKAVISLDNPSAVI 135
DB      178 KAEHLKPGAGERLKGALDAYF-----FVGYPFGAISLSELAISGLVPSIGPEADK 231
QY      136 VLDK-----EDI--AVLYPDKTYNTSI---WVPG--EPDKIIVNETK 173
DB      232 ILEKYSFFSKDVVPAGAYKDVAEPTTLAVAAQWVTSKQPDLLI-YNITK 280

RESULT 15
S74B12
ABC-type transport protein sl11623 - Synecchocystis sp. (strain PCC 6803)
N/Alternate names: protein sl11623
C/Species: Synecchocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
C/Accession: S74B12
R/Kaneko, T.; Sato, S.; Kori, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
S.
A/Reference number: S74322, MUID: 97061201, PMID: 8905231
A/Accession: S74B12
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-223 <KAN>
A/Cross-references: EMBL:D30909; GB:AB001339; NID:gl652844; PIDN:BAAL7773.1; PID:dl01850
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C/Keywords: ATP; nucleotide binding; P-loop; transport protein
F/22-202/Domain: ATP-binding cassette homology <ABC>
F/39-46/Region: nucleotide-binding motif A (P-loop)

Query Match      8.1%, Score 84.5, DB 2, Length 223,
Best Local Similarity 22.5%, Pred. No. 12;
Matches 42; Conservative 33; Mismatches 61; Indels 51; Gaps 9;

QY      18 LALIAGFATQSPINSPYATGTAQAVSEPI-----DVESHILGSTPAAGAGQSGDDIGYAI 72
DB      49 LRLLAGLLTPQSL-----VKLAEPMGFVFQNPEDHQL--VMPTVGA----DIAFGL 93
QY      73 VTIKQVNDVKLVTLRNEOL-----KPYFKYLIQITSGYETNSTALGNFSETKA 124
DB      94 V--KEKLSPOEVQVRVGEALQAVNLAEMARRPIY-----ALSGGQKORIAIAGAIARHCS 146
QY      125 VISLDNPSAVILVDKEDIADVLYPDKTYNTN--TSIWWPG-----EPDKIIV 168
DB      147 VLLDDEFTALDRDSQQELVQVQLVQKRGITATLWVTHRLDELDDYCDGAFLEEGKIVA 206
QY      169 YNETKPV 175
DB      207 QGDEKPL 213
```

Search completed: April 6, 2004, 12:41:49
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:34:02 ; Search time 17 Seconds

(without alignments)
634.031 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTLIAIGIISAAALAL.....EGMLFDSLPIVFQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95.5	9.2	262	1	PLC1 CAEEL
2	91	8.7	427	1	Q27549 methanobact
3	88.5	8.5	334	1	FLA3 PYRHO
4	86.5	8.3	284	1	STOM MOUSE
5	86.5	8.3	846	1	SP98 YEAST
6	86	8.3	1258	1	ICEN ERWHE
7	85.5	8.2	329	1	BCSP BRUME
8	83.5	8.0	468	1	VGLC HSEVB
9	83.5	8.0	870	1	RPAL METH
10	83	8.0	1247	1	NIDO HUMAN
11	83	8.0	1322	1	ICEA PANAN
12	82.5	7.9	243	1	AXI7 ARATH
13	82.5	7.9	236	1	KHSE_CLOAB
14	82.5	7.9	633	1	GIA2 FUSNB
15	82.5	7.9	1122	1	DSG2 MOUSE
16	82	7.9	350	1	OTC RANCA
17	81.5	7.8	417	1	SYS LEPIB
18	81.5	7.8	493	1	FTSA HELPJ
19	81.5	7.8	663	1	RGPI YEAST
20	81.5	7.8	712	1	MUTB RHIME
21	81	7.8	611	1	GIDA MYCFU
22	80.5	7.7	161	1	FMAX_BACNO
23	80.5	7.7	2104	1	MY53_SCHPO
24	80	7.7	196	1	YB17_MERJA
25	80	7.7	537	1	CH61_COREF
26	80	7.7	612	1	GIDA MYCFU
27	80	7.7	880	1	LYTD_BAGSU
28	79.5	7.6	158	1	FMAA_BACNO
29	79.5	7.6	538	1	CAT1_CLOKL
30	79.5	7.6	749	1	VP4_ROTGA
31	79.5	7.6	1196	1	XPG_XENLA
32	79.5	7.6	1461	1	NEOI HUMAN
33	79.5	7.6	2432	1	Y43R_IRV6

34 Q9jmb5 rattus norv
35 P16480 Bacillus th
36 Q9jkl1 mus musculu
37 P39099 escherichia
38 Q9j502 fowlpox vir
39 Q17778 caenorhabdi
40 P40583 saccharomyc
41 P46951 saccharomyc
42 P09846 methanobact
43 Q04544 methamptoc
44 Q60384 methanococc
45 P73141 synechocyst

79 7.6 407 1 ADRM_RAT
79 1180 1 C4AA_BACTI
78 7.5 407 1 ADRM_MOUSE
78 455 1 DEQO_ECOLI
78 512 1 V233_FOWPV
77.5 7.4 375 1 YKUS_CAEEL
77.5 7.4 537 1 YIV9_YEAST
77.5 7.4 817 1 YG4B_YEAST
77.5 7.4 865 1 RPAI_METTW
77.5 7.4 1788 1 POLN_SOUV3
77 7.4 382 1 Y077_METJA
77 7.4 404 1 SYV_SYNV3

ALIGNMENTS

RESULT 1

PLC1 CAEEL STANDARD; PRT; 262 AA.
AC Q93841; Q93783;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative 1-acyl-sn-glycerol-3-phosphate acyltransferase F59F4.4
DE (EC 2.3.1.51) (1-AGP acyltransferase) (1-AGPAT) (lysophosphatidic acid acyltransferase) (LPAAT).
GN F59F4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Harris B., Dobson R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Converts lysophosphatidic acid (LPA) into phosphatidic acid by incorporating acyl moiety at the 2 position (By similarity).
CC -1- CATALYTIC ACTIVITY: Acyl-CoA + 1-acyl-sn-glycerol 3-phosphate = CoA + 1,2-diacyl-sn-glycerol 3-phosphate.
CC -1- PATHWAY: De novo phospholipid biosynthesis; second step.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the 1-acyl-sn-glycerol-3-phosphate acyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z81095; CAB03160.1;
DR EMBL; Z81089; CAB03160.1; JOINED.
DR EMBL; Z81089; CAB03139.1;
DR EMBL; Z81095; CAB03139.1; JOINED.
DR PIR; T22599; T22599.
DR WormPep; F59F4.4; CE11552.
DR InterPro; IPR002123; Acyltransferase.
DR InterPro; IPR004552; AGP acyltrn.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; PlsC; 1.
DR TIGRPFAMs; TIGR00530; AGP acyltrn; 1.
KW Hypothetical protein; Phospholipid biosynthesis; Transferase;
KW Acyltransferase; Transmembrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 89 109 POTENTIAL.
SQ SEQUENCE 262 AA; 29638 MW; 0361FE6C9710593E CRC64;

Query Match 9.2%; Score 95.5; DB 1; Length 262;
Best Local Similarity 24.2%; Pred. No. 0.72;
Matches 58; Conservative 27; Mismatches 96; Indels 59; Gaps 9;

QY 4 TLLAAGIISAAALA-----LLAGPATT---OSPLNSFYATG 38
DB 2 TFLAFLVIAVILLAAQLPVIGFIRAVYFGMCLIIIGFLLGLASIPFGSPNNHFMFK 61
QY 39 TAQAVSEPIDVESHLS-----ITPAAGAGQSDDIGYAIWTKQVNDVKLVKLT 88
DB 62 IFQAMTWPMGVRELRNSEILHDKKPYIIIANHQSALDVLGMSFAPVDCV---VMLKSSL 119
QY 89 RNAEOLKPYKYLIQITSGYETNSTALGNFSEKAVISLDNPSAVILDKEDIAVLDPD 148
DB 120 -----KYLPGFNLCAYLSDSVYINRFSKEALKUTVDTLHEIVTKRKWMI-YPE 168
QY 149 KTGYNSTSIWPGEPDKIIVNETK---PVAIIINFKAYEAK-----GMIFDSLPVI 198
DB 169 GTRNAEPFL-LPFPKGFALAKQAKIPIVPCVFSHKKFFYSHAEKRLTSGNCIIIDILPEV 227

RESULT 2
YF05 METH
ID YF05 METH STANDARD; PRT; 427 AA.
AC O27549;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Hypothetical protein MTH1505.
GN MTH1505.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDouall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- SIMILARITY: Belongs to the ATZ/TRZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000910; AAB85980.1; -.
DR PIR; G69067; G69067.
DR InterPro; IPR006680; Amidohydro 1.
DR Pfam; PF01979; Amidohydro_1; 1.
DR ProDom; PD000518; Urease; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 427 AA; 46441 MW; 439527DD142182CD CRC64;

Query Match 8.7%; Score 91; DB 1; Length 427;
Best Local Similarity 22.6%; Pred. No. 3.1;
Matches 40; Conservative 29; Mismatches 80; Indels 28; Gaps 6;

QY 22 AGFATTQSPNSFYATGTAAQAVSEPIDVESHLSITPA-----AGAAGSDDIGYAIWTK 76

DB 197 AALADKNDLMIHIVHSETENEYSE-----VSRSHGMTPEVYLDDEVGLGPRTPVAACHCVWLK 252
QY 77 DQVNDV-----KUKVT-----LRNAEQLKPYKYLIQITSGYETNSTALGN-----FSET 122
DB 253 DWEIDVLAERDVKVSNNKSLASGVSPVALLQRGVNVSLGTGGAASNNLDFQEM 312
QY 123 KAV-----ISLDNPSAVILDKEDIAVLDPDKTGYNTSISWPGEPDKIIVNETKP 174
DB 313 KTASLLQKNVLEDPTALPMDVFSMATLNGARALGIDAGLIAPKGLADIVILNTRRP 369

RESULT 3
FLA3 PYRHO
ID FLA3 PYRHO STANDARD; PRT; 334 AA.
AC O58284;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Probable flagellin PH0549.
GN PH0549.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- FUNCTION: Flagellin is the subunit protein which polymerizes to
CC form the filaments of archaeal flagella (By similarity).
CC -!- SIMILARITY: Belongs to the archaeal flagellin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000002; BAA29638.1; -.
DR PIR; A71169; A71169.
DR InterPro; IPR002774; Arch_flagellin.
DR Pfam; PF01917; Arch_flagellin; 1.
KW Flagellum; Multigene family; Complete proteome.
SQ SEQUENCE 334 AA; 35438 MW; 9CEAAA6E07C92296 CRC64;

Query Match 8.5%; Score 88.5; DB 1; Length 334;
Best Local Similarity 22.1%; Pred. No. 3.7;
Matches 42; Conservative 30; Mismatches 63; Indels 55; Gaps 8;

QY 8 IAGIISAAALALL-----AGFATTQSPNSFYAT---GTAQAVSEPIDVES 51
DB 15 IAWLVAAVAAAVLINTSGVYLQKQSQTGRQTQTVASGIKVTWVGKADSATPTVIOE 74
QY 52 HLGSITPAAGQSDDIGYAIWTKQVNDVKLVNABOLKPYKYLIQITSGYET 111
DB 75 LAVYITPNAGSSG-----IDLTKVITLSDG-QKQAIKFY----- 108
QY 112 NSTALGNFSEPKAVIS--LDNPSAVILDKEDIAVLDPDKTGYNTSISWPGEPDKIIV 169
DB 109 ---RVGSANELYFIAELMQNLGATIKFDNGSVQV-YFDPDWTISAA-----PTVIID 157
QY 170 NETRPVAILN 179

RA Mayfield J.E., Bricker B.J., Godfrey H., Crosby R.M., Knight D.J.,
RA Halling S.M., Balinsky D., Tabatabai L.B.;
RT "The cloning, expression, and nucleotide sequence of a gene coding
RT for an immunogenic Brucella abortus protein.";
RL Gene 63:1-9(1988).
CC -!- MISCELLANEOUS: BRUCELLA ABORTUS IS THE CAUSATIVE AGENT FOR
CC BRUCELLOSIS IN CATTLE AND MAN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AR009521; AAL51977.1; -;
DR EMBL; M20404; AAA22993.1; -;
DR PIR; AF3351; AF3351.
DR PIR; J02079; IMKB8B.
KW Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 329 31 kDa IMMUNOGENIC PROTEIN.
SQ SEQUENCE 329 AA; 34273 MW; 4EPA2330D0A2544E CRC64;

Query Match 8.2%; Score 85.5; DB 1; Length 329;
Best Local Similarity 19.3%; Pred. No. 6.5; Mismatches 73; Indels 133; Gaps 13;
Matches 56; Conservative 28;

QY 2 KYTLAAGIAGIAAALALLAGPATTQSPINLSPFYATGTAQVASEPDIIVSHLSITPAAG 61
Db 6 KIRLAVAAV---AGAILGASFAVAQAP--TFRIGTGCTAGTYPIGGLIANALSGAG 60
QY 62 AQG-----SDDIGVAIWI-----KQVND 81
Db 61 EKGVPGLVATVSSGVANAINAKSGALESGTQSDVAY---WAYNGTGLYDGKGVED 117
QY 82 VKLKVTIR----- 89
Db 118 LRLALTYPTHTHVARKDANIKSVADLKGKRVSLDEPGSGTIVDARIVLEAYGLTDDI 177
QY 90 NARQLKP-----YFKYLQIQTSGYETNSTALGNFSETKKAVISLDNPSAVI 135
Db 178 KAEHLKPGAGERLKDGLDAYF-----FVGGYPTCAISELAINSGISLVPISGPEADK 231
QY 136 VLDK-----EDI--AVLPDKTYNTSI---WVPG--BDKLIIVNETK 173
Db 232 ILEKYSFFSKDVVPAGAYKDVETPTLAVAAQWTSKAPQDDLI--YNITK 280

RESULT 8
VGLC_HSVB STANDARD; PRT; 468 AA.
AC P12889; P36321;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycoprotein C precursor (Glycoprotein 13).
GN GC OR GP13 OR 16.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1), and
OS Equine herpesvirus type 1 (strain Kentucky D) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520, 10330;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB4P;
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=Kentucky D;
RX MEDLINE=88275055; PubMed=2455821;
RA Allen G.P., Coogale L.D.;
RT "Characterization of an equine herpesvirus type 1 gene encoding a
RT glycoprotein (gp13) with homology to herpes simplex virus
RT glycoprotein C.";
RL J. Virol. 62:2850-2858(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Kentucky D;
RX MEDLINE=89382761; PubMed=2550665;
RA Guo P., Goebel S., Davis S., Perkus M.E., Languet B., Desmettre P.,
RA Allen G., Paoletti E.;
RT "Expression in recombinant vaccinia virus of the equine herpesvirus 1
RT gene encoding glycoprotein gp13 and protection of immunized
RT animals.";
RL J. Virol. 63:4189-4198(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Kentucky D;
RX MEDLINE=93212524; PubMed=8384760;
RA Matsumura T., Smith R.H., O'Callaghan D.J.;
RT "DNA sequence and transcriptional analyses of the region of the
RT equine herpesvirus type 1 Kentucky A strain genome encoding
RT glycoprotein C.";
RL Virology 193:910-923(1993).
CC -!- SIMILARITY: Belongs to the herpesviruses glycoprotein C family.
CC -!- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L07272; AAA46078.1; -;
DR EMBL; M86664; AAB02451.1; -;
DR EMBL; M19966; AAA46077.1; -;
DR EMBL; M29234; AAA46085.1; -;
DR EMBL; S57839; AAB25944.1; -;
DR PIR; A28149; VGBE8H.
DR PIR; A46114; B46114.
DR InterPro; IPR001038; Gp13_EHV.
DR InterPro; IPR00110; Ig-like.
DR InterPro; IPR001654; Marek A.
DR Pfam; PF02124; Marek A; 1.
DR PRINTS; PR00668; GLYCOPROTEIN C.
DR PROSITE; PS50835; IG LIKE; 2.
KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal; Repeat.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 468 GLYCOPROTEIN C.
FT TRANSMEM 432 451 POTENTIAL.
FT DOMAIN 220 311 IG-LIKE 1.
FT DOMAIN 321 416 IG-LIKE 2.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 62 62 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 107 107 E -> K (IN REF. 4).
FT CONFLICT 145 145 E -> K (IN REF. 4).
FT CONFLICT 275 275 V -> A (IN REF. 4).
SQ SEQUENCE 468 AA; 50889 MW; 86F67AFD15AF1C89 CRC64;

Query Match 8.0%; Score 83.5; DB 1; Length 468;
Best Local Similarity 22.5%; Pred. No. 15;
Matches 46; Conservative 33; Mismatches 72; Indels 53; Gaps 9;

```
QY 6 LAIAGIISAALALLAGFAATTSQPLNSFYATGTAQAVSEPIDVESHLSGITPAAGAGQS 65
Db 10 VAVAYLICAGALLTYASGASASSSQ-----SIPATHTHTPNTLTTAHGA-GS 55
QY 66 DDTGVAIVWKQDVNDVKLVTLNRAEQLPKPYKYLQIQTSGYETNSTALG-NFSETKA 124
Db 56 DNTTANG--TESTHSHETTITCTKSLISVPYKQVDMNCT-----TSVGVNYSYRL 106
QY 125 VISLNDPSAVILDKEDIAVLDPKGTNTSIWVGPEDKIIYVNETKPVAILNPKAFY 184
Db 107 EIVLN-----ORTPESGTP-----PGDEENYINHNATKQDTLLLFSTAE 145
QY 185 EAKE-----GMLPDSLP--VIFN 200
Db 146 RKKSRRGGQLGVIPDRLPKROLFN 169

RESULT 9
RPA1 METH
ID RPA1 METH STANDARD; PRT; 870 AA.
AC 027125;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6).
GN RPOA1 OR MTH1051.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=93711463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniel C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA delah: functional analysis and comparative genomics.";
RT J. Bacteriol. 179:7135-7155 (1997).
RL CC
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- COFACTOR: Zinc.
CC -!- SUBUNIT: M.thermoautotrophicum RNA polymerase is composed of four
CC subunits: A', A'', B, and B'.
CC -!- SIMILARITY: THE COMBINED A'+A" SUBUNITS CORRESPOND TO THE A
CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
CC EUKARYOTIC BETA' SUBUNIT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000876; AAB85542.1; --
CC PIR; G69006; G69006.
CC InterPro; IPR000722; RNA_pol_A.
CC InterPro; IPR007080; RNA_pol_Rpb1_1.
CC InterPro; IPR007066; RNA_pol_Rpb1_3.
CC InterPro; IPR007083; RNA_pol_Rpb1_4.
CC InterPro; IPR007081; RNA_pol_Rpb1_5.
CC InterPro; IPR006592; RNA_pol_A_N.

DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR SMART; SM00663; RPOA_N; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW Complete proteome; Zinc-finger.
FT ZN FING 60 103 C4-TYPE (POTENTIAL).
SQ SEQUENCE 870 AA; 98077 MW; 05BB2C280CFD211 CRC64;

Query Match 8.0%; Score 83.5; DB 1; Length 870;
Best Local Similarity 21.7%; Pred. No. 31;
Matches 46; Conservative 32; Mismatches 73; Indels 61; Gaps 9;

QY 46 PIDVESHLSGITPAAGAGSDSDIGYAVIWMKQDVNDVKLVTLR-NAEQI--KPYKYLQ 102
Db 212 PVTVRP---SITLETGSEDDLTHTKLVLDL-RINQ-RUKENMEAGAPQLIVEDLWELLQ 266
QY 103 IQTSGYETNSTALG-----NFSETKAV----- 125
Db 267 YHVTTFVDNEASGVPPARHRSGRPLKTLAQLKKGEGFRFNSLGRKVNFSARTVISDP 326
QY 126 -ISLNDPSAVILDKEDIAVLDPKGTNTSIWVGPED-----KIIVYN 170
Db 327 NISINEVGVEPIIAREVTVPVYVTEWNIDRMREYIENGPDVPHGNAVIVIRPDGKIRIYN 386
QY 171 ETKPVAILNFKAFYEAKEGMLFDSLPIVFNQ 202
Db 387 ETRKVEVLNFKPGY-IVERHLKDGDIVLFNRQ 417

RESULT 10
NIDO HUMAN
ID NIDO HUMAN STANDARD; PRT; 1247 AA.
AC P14543; Q14942;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nidogen precursor (Entactin).
GN NID.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90091745; PubMed=2574658;
RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
RA Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
RA "Human nidogen: complete amino acid sequence and structural domains
RA deduced from cDNAs, and evidence for polymorphism of the gene.";
RL DNA 8:581-594 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=96044428; PubMed=7557988;
RA Zimmermann K., Hoischen S., Hafner M., Nischt R.;
RA "Genomic sequences and structural organization of the human nidogen
RA gene (NID).";
RN [3]
RP SEQUENCE OF 667-1247 FROM N.A.
RX TISSUS=Placenta;
RA MEDLINE=89270475; PubMed=2471408;
RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,
RA Timpl R., Chu M.-L., Uitto J.;
RA "Human nidogen: cDNA cloning, cellular expression, and mapping of the
RA gene to chromosome 1q43.";
RL Am. J. Hum. Genet. 44:876-885 (1989).
CC -!- FUNCTION: Sulfated glycoprotein which is widely distributed in
CC basement membranes and that is tightly associated with laminin.
CC Also binds to collagen IV. It probably has a role in cell-
CC extracellular matrix interactions.
```

CC -!- SUBUNIT: Interacts with PBLN1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Basement membranes.
 CC -!- PTM: N- and O-glycosylated.
 CC -!- SIMILARITY: Contains 6 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 thyroglobulin type-I domain.
 CC -!- SIMILARITY: Contains 5 LDL-receptor YWTD domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M30269; AAA59932.1; -;
 DR EMBL; X82245; CAA57709.1; -;
 DR EMBL; X84819; CAA57709.1; JOINED.
 DR EMBL; X84820; CAA57709.1; JOINED.
 DR EMBL; X84821; CAA57709.1; JOINED.
 DR EMBL; X84822; CAA57709.1; JOINED.
 DR EMBL; X84823; CAA57709.1; JOINED.
 DR EMBL; X84824; CAA57709.1; JOINED.
 DR EMBL; X84825; CAA57709.1; JOINED.
 DR EMBL; X84826; CAA57709.1; JOINED.
 DR EMBL; X84827; CAA57709.1; JOINED.
 DR EMBL; X84828; CAA57709.1; JOINED.
 DR EMBL; X84829; CAA57709.1; JOINED.
 DR EMBL; X84830; CAA57709.1; JOINED.
 DR EMBL; X84831; CAA57709.1; JOINED.
 DR EMBL; X84832; CAA57709.1; JOINED.
 DR EMBL; X84833; CAA57709.1; JOINED.
 DR EMBL; X84834; CAA57709.1; JOINED.
 DR EMBL; X84835; CAA57709.1; JOINED.
 DR EMBL; X84836; CAA57709.1; JOINED.
 DR EMBL; X84837; CAA57709.1; JOINED.
 DR EMBL; M27445; AAA57261.1; -;
 DR PIR; A33322; MMHUND.
 DR PDB; 1NDX; 29-DEC-99.
 DR Genew; HGNC:7821; NID.
 DR MIM; 131390; -;
 DR InterPro; IPR000152; Asx_hydroxyl_5.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006605; G2F_Like.
 DR InterPro; IPR009017; G2F_Like.
 DR InterPro; IPR000033; Ldl_receptor_rep.
 DR InterPro; IPR003886; Nidogen_ext.
 DR InterPro; IPR00716; Thyroglobulin_1.
 DR Pfam; PF00008; EGF; 6.
 DR Pfam; PF00058; ldl_recept_b; 3.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00682; G2F; 1.
 DR SMART; SM00135; LY; 5.
 DR SMART; SM00539; NIDO; 1.
 DR SMART; SM00211; TY; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS00026; EGF_3; 5.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR PROSITE; PS00484; THYROGLOBULIN_1; 1.
 KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;
 KW Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion;
 KW 3D-structure.
 FT CHAIN 1 28
 FT SIGNAL 29 1247
 FT NIDOGN.
 FT DOMAIN 29 669
 FT I (LARGER GLOBULAR DOMAIN).
 FT DOMAIN 670 917
 FT I (CYSTEINE-RICH).
 FT DOMAIN 918 1247
 FT III (SMALLER GLOBULAR DOMAIN).
 FT DOMAIN 386 426
 FT EGF-LIKE 1.
 FT DOMAIN 668 709
 FT EGF-LIKE 2.

FT DOMAIN 710 751 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 758 801 EGF-LIKE 4.
 FT DOMAIN 802 840 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 872 919 THYROGLOBULIN TYPE-I.
 FT DOMAIN 989 1030 LDL-RECEPTOR YWTD MOTIF 1.
 FT DOMAIN 1032 1073 LDL-RECEPTOR YWTD MOTIF 2.
 FT DOMAIN 1075 1118 LDL-RECEPTOR YWTD MOTIF 3.
 FT DOMAIN 1124 1163 LDL-RECEPTOR YWTD MOTIF 4.
 FT DOMAIN 1208 1244 EGF-LIKE 6.
 FT MOD_RES 289 289 SULFATION (POTENTIAL).
 FT MOD_RES 296 296 SULFATION (POTENTIAL).
 FT DISULFID 672 685 BY SIMILARITY.
 FT DISULFID 679 695 BY SIMILARITY.
 FT DISULFID 697 708 BY SIMILARITY.
 FT DISULFID 714 727 BY SIMILARITY.
 FT DISULFID 721 736 BY SIMILARITY.
 FT DISULFID 738 750 BY SIMILARITY.
 FT DISULFID 762 777 BY SIMILARITY.
 FT DISULFID 769 787 BY SIMILARITY.
 FT DISULFID 789 800 BY SIMILARITY.
 FT DISULFID 806 817 BY SIMILARITY.
 FT DISULFID 811 826 BY SIMILARITY.
 FT DISULFID 828 839 BY SIMILARITY.
 FT DISULFID 1212 1223 BY SIMILARITY.
 FT DISULFID 1219 1232 BY SIMILARITY.
 FT DISULFID 1234 1243 BY SIMILARITY.
 FT SITE 702 704 CELL ATTACHMENT SITE.
 FT CARBOHYD 1137 1137 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 33 34 EL -> SS (IN REF. 2).
 FT CONFLICT 37 42 FGFQGG -> SAPDR (IN REF. 2).
 FT CONFLICT 1115 1115 H -> T (IN REF. 3).
 FT SEQUENCE 1247 AA; 136488 MW; 468185B3CEC1575B CRC64;
 Query Match 8.0%; Score 83; DB 1; Length 1247;
 Best Local Similarity 23.7%; Pred. No. 52;
 Matches 47; Conservative 23; Mismatches 72; Indels 56; Gaps 9;
 QY 28 QSPINSFYATGTA-QAVSEPIDVESHLSITPAAGA-----QGSDDIGYALVWKQ 78
 DB 68 RSDIDAVYVTNGIIATSEPPAKESHPLFPPTFGAVAPFLADLDTDGUGK--VYTR- 124
 QY 79 VNDVKLVTLRNAEQKPKYFKYLIQITSGYETNSTALGNFSETKAVISLDNFSAVIVLD 138
 DB 125 --DLSPSITQRAAECHVRGPEISFQ-----PSSAVVVT 156
 QY 139 KEDIAVLY-----PDKTYNTSITWPGEDPK-----IIVNETKPVAILNFKAFYEAKG 189
 DB 157 WESVAPYQGESRDPDQKGKENTFQAVLASDSSSYAIFLYPEDG---LQFHTTFSKKEN 212
 QY 190 MLFDSLPLVIEFQVLQVG 207
 DB 213 ---NQVPAAVAFSGSGVG 227
 RESULT 11
 ICEA PANAN
 ID ICEA PANAN STANDARD; PRT; 1322 AA.
 AC P20469;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inaA.
 GN INAA.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 CX NCBI_TaxID=553;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90092494; PubMed=2599095;
 RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
 RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
 to those of Pseudomonas species and regions required for ice

```

RT nucleation activity."
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17316; CAA35194.1; -.
CC PIR; S07053; S07053.
CC HSP; P06620; IINA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleatn; 69.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 49.
CC Ice_nucleation; Repeat; Outer membrane.
CC DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;
CC
CC Query Match 8.0%; Score 83; DB 1; Length 1322;
CC Best Local Similarity 23.2%; Pred. No. 56;
CC Matches 46; Conservative 27; Mismatches 65; Indels 60; Gaps 11;
CC
CC 3 YTTLAIAGIIASAALALLAGFATQTSP-LNSFYAT--GTAQAVSEPIDVESHLGSIIPA 59
CC 644 YGSTSTAG-----ADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQESDLTAGVGS-TST 697
CC 60 AGAQGSDDIGYAIWIKDQNDVVKLVTLNNAEQLEKPYFYL-----QIQITS 107
CC 698 AGADSSLIAGY-----GSTQTASYHSSLTAGYGSTQTAQEQSVLIT 738
CC 108 GYETNSTA-----LGNFSETK-----AVISLDNPSAVIILDKEDIAVLDPDKTYNTS 156
CC 739 GYGSTSTAGADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQERSDLT-----TGYGSTS 792
CC
CC 157 IWVGPEDPKIIV-YNETK 173
CC 793 --TAGADSSLIAGYGSTQ 808
CC
CC RESULT 12
CC ID AX17 ARATH STANDARD; PRT; 243 AA.
CC AC Q38825;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Auxin-responsive protein IAA7 (Indoleacetic acid-induced protein 7).
CC GN IAA7 OR AT3G23050 OR MXC7.8.
CC OS Arabidopsis thaliana (Mouse-ear cress).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
CC OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CC OK NCBI_taxid=3702;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=cv. Columbia;
CC EX MEDLINE=95387393; PubMed=7658471;
CC RA Abel S., Nguyen M.D., Theologis A.;
CC RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in
CC Arabidopsis thaliana."
CC J. Mol. Biol. 251:533-549(1995).

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.,
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.",
RL DNA Res. 7:131-135(2000).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Barn J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vayenberg M., Wallander E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.",
RL Science 302:842-846(2003).
CC -!- FUNCTION: Could act as regulator of genes responsible for
CC mediating the various auxin-induced events responsible for cell
CC growth (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- INDUCTION: By auxin.
CC -!- SIMILARITY: Belongs to the AUX/IAA family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18409; AAC49048.1; -.
CC EMBL; AB026655; BAB02096.1; -.
CC EMBL; AF323395; AAG48759.1; -.
CC PIR; S58494; S58494.
CC TRAFAC; T04536; -.
CC InterPro; IPR003311; AUX_IAA.
CC Pfam; PF02309; AUX_IAA, I.
CC Multigene family; Nuclear protein; Translation regulation.
CC SEQUENCE 243 AA; 26381 MW; 8305B9EA2C2B5D0 CRC64;
CC
CC Query Match 7.9%; Score 82.5; DB 1; Length 243;
CC Best Local Similarity 25.0%; Pred. No. 8;
CC Matches 54; Conservative 34; Mismatches 79; Indels 49; Gaps 12;
CC
CC 13 ASAAALALLAGFATQTSPNSFYATQAVSEPIDVESHLGSIIPAAGQSDDIGYAI 72
CC 10 ATELCLGLPGGAEEVSEPAKS--AVGSKRGFSETVDMLMLNLS-----NKEGSVD----- 57
CC 73 VLIKQNDVVKLVTLNNAEQ-----LKPFFKYLQIQ--ITSG-YETNSTALG 117
CC 58 --LKNVSAVPKTKTLKPSKPPAKAQVGVGPPVRYRNKMMTQTKTSSAEASSEKAG 115
CC 118 NFSETKA-----VISLDNPSAVIILDKEDIAVL--YPDKTG-----YTNTSIWVGPEDPK 165
CC 116 NFGGAGAGAGLVKVSMDGAP---YLKVDLKMYSQYQDLSDALAKMFSSTMGNYGAQGM 172
CC 166 IIVYNETKPVAILNFKAF---YEAKEG--MLFDSLFP 196
CC 173 IDFMNESKLMNLLNSSEYVPSYEDKGDGMVLVGDPV 208

```


[illegible]

Search completed: April 6, 2004, 12:40:17
Job time : 18 secs


```

118 NFSETKAVIS---LDNPSAVILD--KEDIAVLDPKTYGNTSITWVPGEDKII-VYNE 171
      :      :      :      :      :      :      :      :      :      :
250 QI-ORT-EVPT-GEFEAHDPSQPIVIAALKGVTSALGDRINTYNASLEAKGGIRLVKDE 409
      :      :      :      :      :      :      :      :      :      :

```

Qy	172	T	172
D _b	410	T	410

RESULT 2

```

Q53651      PRELIMINARY;          PRT;   255 AA.
ID          Q53651
AC          Q53651;
DT          01-NOV-1996 (TrEMBLrel. 01, Created)
DT          01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT          01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE          cAMP factor precursor.
DE          OS
OS          Streptococcus agalactiae.
OC          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC          Streptococcus.
OX          NCBI_TaxID=1311;
OX          [1]
RN          SEQUENCE FROM N.A.
RP          STRAIN=R268;
RC          RX
RC          MEDLINE=95231384; PubMed=7715536;
RA          podbielski A.;
RA          RT
RT          "Molecular characterization of the cf6 gene encoding gro
RL          streptococcal CAMP-factor."
RL          Med. Microbiol. Immunol. 183:239-256 (1994).
DR          EMBL; X72754; CAA51283.1; -.
KW          Signal; CAMP.
FT          SIGNAL          1      29      POTENTIAL.
FT          CHAIN          30     255      POTENTIAL.
SO          SEQUENCE 255 AA; 283373 MW; 5061CE85C31DF894 CRC64;

```

Query Match 9.2%; Score 96; DB 2; Length 255;
Best Local Similarity 25.0%; Pred. No. 4.5;
Matches 53; Conservative 31; Mismatches 84; Indels

QY	5	TLIATIGIISAAALALAGFATTQSPINFSVFATGTAQAVSEPIDVES--HLGSITPAAGAQ	63
Db	12	TLVAGALLFSPAVLEVHADQVTTTPQVNVHNSNQAOQMAQLQDQSILQENIKD--NVQ	69
QY	64	GSD---DIGYALVWIKQDVNDVKLVLR-----NAEQLKPYFKYLIQIT	106
Db	70	GTDYEKPVNEAITSVE-----KLTSURANPETYDLNIGSRVEALTDVIE--AITFS	121
QY	107	SGYETNSTALGN---FSETKAVISLDNPSAVI-----VLDKEDIADVLPDKTGYT	153
Db	122	TQHLTNKYSQANI DMGFGITKLIVRLDPFASVDSIKAQVNDVKALEQKVLTPDLKPTD	181
QY	154	NTSIWPGPEDKIIIVNE--TKQVAILNPKAF	183
Db	182	RATITVTKSKLDKEI--WNTFRFTDRDKVLNVKEF	212

DECIU.T 3

Q8VM91	PRELIMINARY;	PRT;	449 AA.
ID Q8VM91			
AC Q8VM91;			
DT 01-MAR-2002	(TrEMBLrel. 20, Created)		
DT 01-MAR-2002	(TrEMBLrel. 20, Last sequence update)		
DT 01-JUN-2003	(TrEMBLrel. 24, Last annotation update)		
DE Proteinase.			
DE PRT.			
OS Lactobacillus helveticus.			
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;			
OC Lactobacillus.			
OX NCBI_TaxID=1587;			
RN [1]			
RC SEQUENCE FROM N.A.			
RC STRAIN=JCM1003;			

RA Saito Y., Hamanaka Y., Takizawa S., Benno Y.;
RT "Rapid Detection of *Lactobacillus helveticus* in Feracal Samples of
RT Healthy Subjects Administered *L. helveticus* Yogurt using S-layer Gene-
RT Targeted Primers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB061776; BAB72066.1; -.
DR GO; GO:0009274; C:cell wall (sensu Bacteria); IEA.
DR GO; GO:0030115; C:S-layer; IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR InterPro; IPR004903; SLAP.
DR Pfam; PF03217; SLAP; 1.
DR PRINTS; PR01729; SURFACEPLAYER.
SO SEQUENCE 449 AA, 47940 MW; EBA765307777D343A CRC64;

Query Match 9.1%: Score 95; DB 2; Length 449;

```

Qy 11 IIAAAALLAGFATTQSPINLFYATGTQAQVSEPIDVESHGCS-----ITFA-- 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 IVSAAALLAVA PVAATAMPVNAATTTTSTTTNKPTVDLSGAGSVSEKDTVNVTSFT 66

```

QY
60
AGRQSDDIGYALVWIKQVNDVKUKVILKRNREQKFIKFIKQ
| | | | : | | | | : | | | : | | | :
-----AGRQSDDIGYALVWIKQVNDVKUKVILKRNREQKFIKFIKQ-----
135

105 ---ITSGVETNSTALG-NFSETKA---VISLNPASAVIVLD-KEDIAVLYPDKTGYTNT 155

Db 126 LSAVEAGKEYTMTLSGVGFSFGKANACKTLLTFKLPKNVKVNDTSNDVKVSLDQYGNATNL 185

[illegible]

RESULT 4

Q80N15	PRELIMINARY; PRT; 1666 AA.
AD	Q80N15;
AC	Q80N15; (TrEMBLrel. 24, Created)
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Structural glycoprotein.
DE	Structural glycoprotein.
OS	Yellow head virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales.
OX	NCBI_TaxID=96029;
OX	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=24541857; PubMed=12655087;
RX	Jitrapakdee S., Unajak S., Sittidilokratna N., Hodgson R.A.J.,
RA	Cowley J.A., Walker P.J., Panyim S., Boonsaeng V.;
RA	"Identification and analysis of gp16 and gp64 structural
RA	glycoproteins of yellow head virus of Penaeus monodon shrimp."
RL	J. Gen. Virol. 84:863-873 (2003).
RL	ENBL; AF540644; AA083987.1; -
DR	SEQUENCE 1666 AA; 195430 MW; C0B523A754D6EA9A CRC64;
SO	SEQUENCE

Query Match	g.0%:	Score 94:	DB 12:	Length 1666:
-------------	-------	-----------	--------	--------------

	Query match	Best Local Similarity	26.48;	Pred. No. 95;	Mismatches	71;	Indels	42;	Gaps	12;
		Matches	51;	Conservative						
QY	20	LLAGFAITQPLNSFYATGTAQAVSEPID--VSSHLSGIITPAAGAQSDDIGYAIWIKD	77							
Db	862	LYLGKVSAAASIGTYSKCGKAQSPSPVHDHGINTDLG--TPVYDS-ACDAAAYTIPVVKY	918							
QY	78	Q-----VNVVKLVTLRNEQLKPYFKYLIQIQTSGVETNSTALGNFSETKAVISLDNP	131							
Db	919	NGPYSLGVPDVSCEI---HDETL-----TCG--TNSTP--RFSICSHKIPYDGP	960							
QY	132	SAVTLVDKEDIAYLPDKTGYNTS IWVGPEDKI -IVYNETKP-----VAILNKA	182							
Db	961	HSVTCISKNKWHVWKPQY---SVYIAGDPGALHISHNKHKEYTILKQDQINLFHFSY	1017							

```

QY 183 FYEKEGMLFDSL 195
Db 1018 LYQA-VAMLFGLS 1029

RESULT 5
Q9LCD3 PRELIMINARY; PRT; 508 AA.
ID Q9LCD3
AC Q9LCD3;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Dipeptide transporter DppA homolog.
GN DPPA.
OS Ochrobactrum anthropi.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Ochrobactrum.
OX NCBI_TaxID=529;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRC SV3;
RX MEDLINE=20193627; PubMed=10727942;
RA Komeda H., Asano Y.;
RT "Gene cloning, nucleotide sequencing, and purification and
RT characterization of the D-sterospecific amino-acid amidase from
RT Ochrobactrum anthropi SV3".
RL Eur. J. Biochem. 267:2028-2035 (2000).
DR EMBL; AB026907; BAA94699.1; -.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
DR SEQUENCE 508 AA; 55661 MW; 84E7579A526CA74E CRC64;

Query Match 8.8%; Score 91.5; DB 2; Length 508;
Best Local Similarity 27.6%; Pred. No. 29;
Matches 40; Conservative 23; Mismatches 47; Indels 35; Gaps 8;

QY 35 YATGTAQAVSEPIDVESHGSGTTPAAGAGSGDDIGYAIWIK--DQVNDVKLVIRNAE 92
Db FSTGT-----PVTVDVIFSLQKAADPQGS--FGFAFEPIKSIDKIDKTVRLTKH-- 147

QY 93 QLKPY-----FKYLIQITSGYETNSTALGNFSETKAVISLONPSAVIVLDKEDIAV 144
Db ---PVSALSAUSLVAAISVSKADYKPSAFG-----SNPVTI--GPFFKVESYERGTOAV 198

QY 145 LYPDKTGYTNTGIWVGEPDKIIVY 169
Db LVP-----NTYYWRQGEDGKALPY 217

RESULT 6
Q8CX23 PRELIMINARY; PRT; 255 AA.
ID Q8CX23
AC Q8CX23;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE cAMP factor.
GN GBS2000.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Meadek T., Zouine M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing

invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766855; CAD47659.1; -.
DR Sagalinst; gbs2000; -.
KW Complete proteome.
SQ SEQUENCE 255 AA; 28374 MW; 266DBB9D0B73562C CRC64;

Query Match 8.7%; Score 91; DB 16; Length 255;
Best Local Similarity 24.5%; Pred. No. 12;
Matches 52; Conservative 31; Mismatches 85; Indels 44; Gaps 10;

QY 5 TLAIAGIIASAAALALAGFATTQGPLNSFYATGTAQAVSEPIDVES-HLGSITPAAGAQ 63
Db TLVAGALLFSPAVLEVHADQVTTTPQVNVHVNNSNQAQMAQLDQDSIQLRNKID--NVQ 69

QY 64 GSD---DIGYAIWIKQVNDVKLVIR-----NAEQLKPYFKYLIQIT 106
Db GTDYKTVNEAITSVE-----KLKTSLRANPEVTYDLSIGSRVEALTDVIE--AITFS 121

QY 107 SGYETNSTALGN---FSETKAVISLONPSAVI-----VLDKEDIAVLPDKTGYT 153
Db TQHLANKVSOANIDMGFGITKLVIRILDPFASVDSIKAQVNDVKALEQKLVTPDLKPTD 181

QY 154 NTSIWVGEPDKIIVYNE--TKPVAILNFKAP 183
Db RATIVTKSKLDKEI-WNTRFTRDKKVLNVKEF 212

RESULT 7
Q8CX01 PRELIMINARY; PRT; 255 AA.
ID Q8CX01
AC Q8CX01;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE CAMP factor.
GN CFB OR SAG2043.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547;
RA Tettelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AF014283; AAN00902.1; -.
DR TIGR; SAG2043; -.
KW Complete proteome.
SQ SEQUENCE 255 AA; 28374 MW; 266DBB9D0B73562C CRC64;

Query Match 8.7%; Score 91; DB 16; Length 255;
Best Local Similarity 24.5%; Pred. No. 12;
Matches 52; Conservative 31; Mismatches 85; Indels 44; Gaps 10;

QY 5 TLAIAGIIASAAALALAGFATTQGPLNSFYATGTAQAVSEPIDVES-HLGSITPAAGAQ 63
Db TLVAGALLFSPAVLEVHADQVTTTPQVNVHVNNSNQAQMAQLDQDSIQLRNKID--NVQ 69

QY 64 GSD---DIGYAIWIKQVNDVKLVIR-----NAEQLKPYFKYLIQIT 106
Db GTDYKTVNEAITSVE-----KLKTSLRANPEVTYDLSIGSRVEALTDVIE--AITFS 121

```

```
QY 107 SGYETNSTALGN-----FSETKAVISLNDNSAVI-----VLDKEDIAVLPDKTGYT 153
Db 122 TQHLANKVSNQIDMGFGITKLVIRILDPFASVDSIKAQVNDVKALEQKVLATYPLDKPTD 181
QY 154 NTSIWPVGPDKIIVNE--TKPVAILNFKAF 183
Db 182 RATIYTKSKLDKEI-WNTRFTRDKKVLNKEF 212

RESULT 8
Q9YCP2 PRELIMINARY; PRT; 533 AA.
ID Q9YCP2
AC Q9YCP2
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein APE1216.
GN APE1216.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococccaeae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101 (1999).
DR EMBL; AP000061; BA880205.1; -.
DR PIR; G72593; G72593.
DR HSP; P04268; IIC2.
DR GO; GO:0005222; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002592; Rec_signal.
DR InterPro; IPR001865; Ribosomal_S2.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF01664; Rec_signal; 1.
DR Pfam; PF00435; Spectrin; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR SMART; SM00150; SPEC; 1.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 533 AA; 56481 MW; E8B53D791C353E31 CRC64;

Query Match 8.7%; Score 91; DB 17; Length 533;
Best Local Similarity 22.4%; Pred. No. 34;
Matches 49; Conservative 24; Mismatches 76; Indels 70; Gaps 9;

QY 5 TLIAAGITASAAALALLAGFATTQSPPLNSFYATGTAAVSPIDVESHLSGITPAAGAGG 64
Db 17 TLIAAPLAGGAGAGQAFVVGSP-----TPVDIVVWLGGEF----- 57

QY 65 SDDIGYAI--VWIKQVNDVKLVTLRNA-----EOLKPYFKYLIQITSGYET 111
Db 58 -----YDVEVWLSGTGVSEVLIALTLGSGEVLGPRSLRPGVYTLRFRALSGL-- 110

QY 112 NSTALGNSETKAVISLNDNSAVI--VLDKE-----DIAVLVDPK 149
Db 111 --TASAVVEVTEATITLDKSGVGVGEVLAETAEVGLGGGVSVAYQLRMGGVTVAVYPDD 168

QY 150 TGYTNTSIWVGPEDKI-----IVYNETKPVAILNFKAFY 184
```

```
Db 169 SGRAVASFTVPLPSPGVHTVELVYT--PPLWLRFSQSY 204

RESULT 9
Q8AV58 PRELIMINARY; PRT; 2169 AA.
ID Q8AV58
AC Q8AV58
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Sidekick-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22217368; PubMed=12230981;
RA Yamagata M., Weiner J.A., Sanes J.R.;
RT "Sidekicks: synaptic adhesion molecules that promote lamina-specific
RT connectivity in the retina.";
RL Cell 110:649-660 (2002).
DR EMBL; AF537107; AAN15075.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004896; F:hematopoietin/interferon-class (D200-domain. . .; IEA.
DR InterPro; IPR000282; Cyto_c_receptor_2.
DR InterPro; IPR003962; FNIII_subd.
DR InterPro; IPR003961; FNIII.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003599; IG_III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR Pfam; PF00041; fn3; 13.
DR Pfam; PF00047; ig; 6.
DR PRINTS; PR00014; FNTPETIIII.
DR SMART; SM00060; FN3; 13.
DR SMART; SM00409; IG; 7.
DR SMART; SM00408; IGC2; 6.
DR PROSITE; PS50835; IG_LIKE; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
SQ SEQUENCE 2169 AA; 239477 MW; 06BFD900AEF19BD7 CRC64;

Query Match 8.7%; Score 91; DB 13; Length 2169;
Best Local Similarity 21.6%; Pred. No. 2.5e+02;
Matches 48; Conservative 34; Mismatches 74; Indels 66; Gaps 9;

QY 2 KYTTLAIGIISAAALALLAGFATTQSPPLNSFYATGTAAVSEPI----- 47
Db 1376 KFTTVEVGSTVROFTATDL-----TPESAVIFRTSAKTQGWGEPLATVITTEKRERPA 1430
QY 48 -----DVESHLSGITPAAGAGSDDIGVAIVWKQVN-----DVKLKVTLR 89
Db 1431 PPQOLTTQSDVSSRSRLHWPVGGSSPIRFTVVQVRELPGNDWQTYSSISHEATSC 1490
QY 90 NAEOLKPYFKY-LQIQITSGYETNSTALGNFS-ETKAVISLND-----PSAVIYLDKEDI 142
Db 1491 IIESLNFTSYKLRVK-----ATNDIGSDYSABTEAVTTLQDVPDPSPSVLV----- 1539
QY 143 AVLYPDKTGYTNTSIWVGPEDKIIVYNETKPVAILNFKAFY 184
Db 1540 -----TPHTTSSVLVQWQPPKA-----ESLNGLLGYRIYY 1570

RESULT 10
Q9I118 PRELIMINARY; PRT; 174 AA.
ID Q9I118
AC Q9I118
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein PA2464.
GN PA2464.
```



```
DR InterPro; IPR004903; SLAP.  
DR Pfam; PF03217; SLAP; 1  
DR PRINIS; PRO1729; SURFACELAYER.  
SQ SEQUENCE 450 AA; 47826 MW; D38D15A9E5B2C13 CRC64;  
  
Query Match      8.6%; Score 89.5; DB 2; Length 450;  
Best Local Similarity 24.0%; Pred.No. 36;  
Matches 52; Conservative 33; Mismatches 87; Indels 45; Gaps 9;  
  
QY    11 IIASAAALLAGPATTQSPLNSFYATGTAQAIVSEPIDVESHLS-----ITPA-- 59  
DB     : ||||| : | : | : | : | : | : | : | : | : | :  
        8 VSA AAAALLAVAPAATAAMPVNAAATVTSTTNKPTVDLGSAGVSSESXOTVVNTPSFT 67  
QY    60 -----AGA QSGDDIGVAIWKIQDVNDVKLVTLRNEAEQLKPYP- - - KYLQ 102  
DB     : ||||| : | : | : | : | : | : | : | : | : | :  
        68 LTS AAKGPATLGSSIEASLNCTSVTADADVAKDVLTDGGKVAVSYDKNTLTNLKSLSD 127  
QY   103 IOITSGYETNSTALGNFSSETKA----VISLDNP S AVIV--- LKDEDIAVI LPDKTGTYN 154  
DB     : : | : | : | : | : | : | : | : | : | :  
        128 VKAGDDYTMTLSGVG-FSF GKANAGKTLPFLPEGTVEGANYNKKDHKVTL-DQYG NVS 185  
QY   155 TSIWVPGE PKLI IVNET KPVAILNFKA FYEAKEGM L 191  
DB     : : | : | : | : | : | : | : | : | : | :  
        186 GLKFV ---ISKVKAYDSANTNAV ---- SFYDAKSGLV 215
```

```
RESULT 15  
Q60045 PRELIMINARY; PRT; 1148 AA.  
ID Q60045; AC Q60045;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Polyalacturonase precursor (EC 3.2.1.82).  
GN PGA.  
OS Thermoanaerobacter thermosulfurogenes (Clostridium  
OC Thermosulfurogenes).  
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
OX Thermoanaerobacteriaceae; Thermoanaerobacterium.  
NCBI_TaxID=33950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EML.  
RA Matsushek M., Sahm K., Bahl H.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.  
CC -- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
      (POLYACTURONASES).  
CC -- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
DR EMBL; U50951; AAC08040.1; -.  
DR PIR; S72635; S72635.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0004650; F:polysaccharonase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPRO03961; FN.III.  
DR InterPro; IPRO08957; FN.III-like.  
DR InterPro; IPRO00743; Glyco_hydro_28.  
DR InterPro; IPRO01119; SLH.  
DR Pfam; PF00041; fn3; 2.  
DR Pfam; PF00295; Glyco_hydro_28; 1.  
DR Pfam; PF00395; SLH; 3.  
DR SMART; SM00060; FN3; 2.  
DR PROSITE; PS00502; POLYACTURONASE; 1.  
DR PROSITE; PS01072; SIH_DOMAIN; 2.  
KW Cell wall; Glycosidase; Hydrolase; Signal.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 1148 POLYACTURONASE.  
SQ SEQUENCE 1148 AA; 122642 MW; E95BD0F2529FA1OE CRC64;
```

Query Match 8.5%; Score 88.5; DB 2; Length 1148;
Best Local Similarity 19.1%; Pred.No. 1.6e+02;
Matches 37; Conservative 38; Mismatches 82; Indels 37; Gaps

Search completed: April 6, 2004, 12:41:16
Job time : 47 secs